

BEST AVAILABLE COPY

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

**CERTIFIED COPY OF
PRIORITY DOCUMENT**

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Stephen Hordley

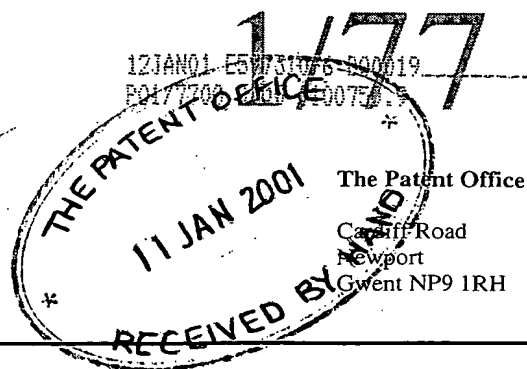
Dated 26 July 2006

THIS PAGE BLANK (USPTO)

The
Patent
Office

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)



1. Your reference P026477GB

2. Patent application number **0100750.9**
(The Patent Office will fill in this part)

3. Full name, address and postcode of the or of each applicant (underline all surnames)

INPHARMATICA LIMITED
60 Charlotte Street
London
W1T 2NU

Patents ADP number (if you know it)

7853286002

If the applicant is a corporate body, give the country/state of its incorporation

United Kingdom

4. Title of the invention NOVEL PROTEINS

5. Name of your agent (if you have one) Carpmaels & Ransford

 "Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode) 43 Bloomsbury Square
London
WC1A 2RA

Patents ADP number (if you know it)

83001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number	Country	Priority application number (if you know it)	Date of filing (day / month / year)
--	---------	---	--

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application	Number of earlier application	Date of filing (day / month / year)
---	-------------------------------	--

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:

a) any applicant named in part 3 is not an inventor, or	Yes
b) there is an inventor who is not named as an applicant, or	
c) any named applicant is a corporate body	

See note (d))

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form

Description 74

Claim(s) 6

Abstract

Drawing(s) 67 267

9

10. If you are also filing any of the following, state how many against each item.

Priority documents

Translations of priority documents

Statement of inventorship and right to grant of a patent (Patents Form 7/77)

Request for preliminary examination and search (Patents Form 9/77)

Request for substantive examination (Patents Form 10/77)

Any other documents (please specify)

11.

I/We request the grant of a patent on the basis of this application.

Signature

Date

Hugh Goodfellow
Carpmaels & Ransford

11th February 2001

12. Name and daytime telephone number of person to contact in the United Kingdom

Hugh Goodfellow

020-7242 8692

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
- If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.
 - Write your answers in capital letters using black ink or you may type them.
 - If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.
 - If you have answered 'Yes' Patents Form 7/77 will need to be filed.
 - Once you have filled in the form you must remember to sign and date it.
 - For details of the fee and ways to pay please contact the Patent Office.

Novel proteins

This invention relates to novel proteins, termed AAC74854.1, AAC76768.1, AAD05645.1, VP4 and Ro60 herein identified as adhesion molecules and to the use of these proteins and nucleic acid sequences from the encoding genes in the diagnosis,
5 prevention and treatment of disease.

All publications, patents and patent applications cited herein are incorporated in full by reference.

BACKGROUND

The process of drug discovery is presently undergoing a fundamental revolution as the era of functional genomics comes of age. The term "functional genomics" applies to an
10 approach utilising bioinformatics tools to ascribe function to protein sequences of interest. Such tools are becoming increasingly necessary as the speed of generation of sequence data is rapidly outpacing the ability of research laboratories to assign functions to these protein sequences.

15 As bioinformatics tools increase in potency and in accuracy, these tools are rapidly replacing the conventional techniques of biochemical characterisation. Indeed, the advanced bioinformatics tools used in identifying the present invention are now capable of outputting results in which a high degree of confidence can be placed.

Various institutions and commercial organisations are examining sequence data as they
20 become available and significant discoveries are being made on an on-going basis. However, there remains a continuing need to identify and characterise further genes and the polypeptides that they encode, as targets for research and for drug discovery.

Recently, a remarkable tool for the evaluation of sequences of unknown function has been developed by the Applicant for the present invention. This tool is a database system,
25 termed the Biopendium search database, that is the subject of co-pending United Kingdom Patent Application No. GB0006153.1. This database system consists of an integrated data resource created using proprietary technology and containing information generated from an all-by-all comparison of all available protein or nucleic acid sequences.

The aim behind the integration of these sequence data from separate data resources is to combine as much data as possible, relating both to the sequences themselves and to information relevant to each sequence, into one integrated resource. All the available data relating to each sequence, including data on the three-dimensional structure of the encoded protein, if this is available, are integrated together to make best use of the information that is known about each sequence and thus to allow the most educated predictions to be made from comparisons of these sequences. The annotation that is generated in the database and which accompanies each sequence entry imparts a biologically relevant context to the sequence information.

10 This data resource has made possible the accurate prediction of protein function from sequence alone. Using conventional technology, this is only possible for proteins that exhibit a high degree of sequence homology (above about 20% homology) to other proteins in the same functional family. Accurate predictions are not possible for proteins that exhibit a very low degree of sequence homology to other related proteins of known function.

15 In the present case, a protein whose sequence is recorded in a publicly available database as AAC74854.1 (NCBI Genbank nucleotide accession number AE000273 and Genbank protein accession number AAC74854.1), is implicated as a novel member of the adhesion molecule family.

20 A second protein, termed AAC76768.1 (NCBI Genbank nucleotide accession number AE000451 and Genbank protein accession number AAC76768.1) is also implicated as a member of the adhesion molecule family.

A third protein termed AAD05645.1 (NCBI Genbank nucleotide accession number AE001445 and Genbank protein accession number AAD05645.1), is implicated as a novel member of the adhesion molecule family.

25 A fourth protein, termed VP4 (NCBI Genbank nucleotide accession number X82323 and Genbank protein accession number CAA57766.1) is also implicated as a member of the adhesion molecule family.

A fifth protein, termed Ro60 (NCBI Genebank nucleotide accession number J04137 and SWISS-PROT protein accession number P10155) is also implicated as a member of the adhesion molecule family.

Adhesion molecules are involved in a range of biological processes, including
 5 embryogenesis (Martin-Bermudo, M.D., *et al.*, Development. 2000 Jun;127(12): 2607-15; Chen, L.M., *et al.*, J Neurosci. 2000 May 15;20(10): 3776-84; Zweegman, S., *et al.*, Exp Hematol. 2000 Apr;28(4): 401-10; Darribere, T., *et al.*, Biol Cell. 2000 Jan;92(1): 5-25), maintenance of tissue integrity (Eckes, B., *et al.*, J Cell Sci. 2000;113(Pt 13): 2455-2462; Buckwalter, J.A., *et al.*, Instr Course Lect. 2000;49: 481-9; Frenette, P.S., *et al.*, J Exp Med.
 10 2000 Apr 17;191(8): 1413-22; Delmas, V., *et al.*, Dev Biol. 1999 Dec 15;216(2): 491-506; Humphries, M.J., *et al.*, Trends Pharmacol Sci. 2000 Jan;21(1): 29-32; Miosge, N., *et al.*, Lab Invest. 1999 Dec;79(12): 1591-9), leukocyte extravasation/inflammation (Lim, L.H., *et al.*, Am J Respir Cell Mol Biol. 2000 Jun;22(6): 693-701; Johnston, B., *et al.*, Microcirculation. 2000 Apr;7(2): 109-18; Mertens, A.V., *et al.*, Clin Exp Allergy. 1993
 15 Oct;23(10): 868-73; Chcialowski, A., *et al.*, Pol Merkuriusz Lek. 2000 Jan;7(43): 13-7; Rojas, A.I., *et al.*, Crit Rev Oral Biol Med. 1999;10(3): 337-58; Marinova-Mutafchieva, L., *et al.*, Arthritis Rheum. 2000 Mar;43(3): 638-44; Vijayan, K.V., *et al.*, J Clin Invest. 2000 Mar;105(6): 793-802; Currie, A.J., *et al.*, J Immunol. 2000 Apr 1;164(7): 3878-86; Rowin, M.E., *et al.*, Inflammation. 2000 Apr;24(2): 157-73; Johnston, B., *et al.*, J Immunol. 2000
 20 Mar 15;164(6): 3337-44; Gerst, J.L., *et al.*, J Neurosci Res. 2000 Mar 1;59(5): 680-4; Kagawa, T.F., *et al.*, Proc Natl Acad Sci U S A. 2000 Feb 29;97(5): 2235-40; Hillan, K.J., *et al.*, Liver. 1999 Dec;19(6): 509-18; Panes, J., 1999 Dec;22(10): 514-24; Arao, T., *et al.*, J Clin Endocrinol Metab. 2000 Jan;85(1): 382-9; Souza, H.S., *et al.*, Gut. 1999 Dec;45(6): 856-63; Grunstein, M.M., *et al.*, Am J Physiol Lung Cell Mol Physiol. 2000 Jun;278(6):
 25 L1154-63; Mertens, A.V., *et al.*, Clin Exp Allergy. 1993 Oct;23(10): 868-73; Berends, C., *et al.*, Clin Exp Allergy. 1993 Nov;23(11): 926-33; Fernvik, E., *et al.*, Inflammation. 2000 Feb;24(1): 73-87; Bocchino, V., *et al.*, J Allergy Clin Immunol. 2000 Jan;105(1 Pt 1): 65-70), oncogenesis (Orr, F.W., *et al.*, Cancer. 2000 Jun;88(S12): 2912-2918; Zeller, W., *et al.*, J Hematother Stem Cell Res. 1999 Oct;8(5): 539-46; Okada, T., *et al.*, Clin Exp Metastasis.
 30 1999;17(7): 623-9; Mateo, V., *et al.*, Nat Med. 1999 Nov;5(11): 1277-84; Yamaguchi, K., *et al.*, J Exp Clin Cancer Res. 2000 Mar;19(1): 113-20; Maeshima, Y., *et al.*, J Biol Chem.

2000 Jun 2 (epub ahead of reprint); Van Waes, C., *et al.*, *Int J Oncol.* 2000 Jun;16(6): 1189-95; Damiano, J.S., *et al.*, *Leuk Lymphoma.* 2000 Jun;38(1-2): 71-81; Seftor, R.E., *et al.*, *Cancer Metastasis Rev.* 1999;18(3): 359-75; Shaw, L.M., *J Mammary Gland Biol Neoplasia.* 1999 Oct;4(4): 367-76; Weyant, M.J., *et al.*, *Clin Cancer Res.* 2000 Mar;6(3): 949-56), and thrombogenesis (Wang, Y.G., *et al.*, *J Physiol (Lond).* 2000 Jul 1;526(Pt 1): 57-68; Matsuno, H., *et al.*, *Nippon Yakurigaku Zasshi.* 2000 Mar;115(3): 143-50; Eliceiri, B.P., *et al.*, *Cancer J Sci Am.* 2000 May;6 Suppl 3: S245-9; von Beckerath, N., *et al.*, *Blood.* 2000 Jun 1;95(11): 3297-301; Topol, E.J., *et al.*, *Am Heart J.* 2000 Jun;139(6): 927-33; Kroll, H., *et al.*, *Thromb Haemost.* 2000 Mar;83(3): 392-6).

- 10 The detailed characterisation of the structure and function of several adhesion-receptor families has led to active programs by a number of pharmaceutical companies to develop adhesion molecule antagonists for use in the treatment of inflammation, oncology, and cardiovascular disease. Adhesion receptors are involved in virtually every aspect of biology from embryogenesis to apoptosis. They are essential to the structural integrity and homeostatic functioning of most tissues. It is therefore not surprising that defects in adhesion receptors cause disease and that many diseases involve modulation of adhesion molecule function.

The Adhesion molecule family in fact represents at least four distinct families which are unified by their function rather than their structure. Of the four families, three are of pharmaceutical interest due to small molecule tractability. They are;

1. The integrin family is a superfamily of α and β heterodimeric transmembrane glycoproteins and is the family, which has attracted most pharmaceutical interest. Its members are large, heavily glycosylated, heterodimeric proteins composed of one of at least 15 distinct α -subunits in non-covalent linkage with one of at least 8 β -subunits. Adhesion receptors bind ligands expressed on cell surfaces, extracellular matrix molecules, and soluble molecules. Integrins are subcategorised based on their β -subunit usage. The members of this family are summarised below in Table 1.
2. Selectins are a small family of three members P, E and L selectin. They are glycoproteins, selectively expressed on cells related to the vasculature, and contain a lectin-binding domain. The members of this family are described below in Table 2.

3. The immunoglobulin family represents the counter receptor for the integrins and includes the intracellular adhesion molecules (ICAMs) and vascular cell adhesion molecules (VCAMs). Members are composed of variable numbers of globular, immunoglobulin-like, extracellular domains. Some members of the family, for example, PECAM-1 (CD31) and NCAM, mediate homotypic adhesion. Some members of the family, for example ICAM-1 and VCAM-1, mediate adhesion via interactions with integrins. The members of this family are described below in Table 3.

Adhesion molecules have been shown to play a role in diverse physiological functions, many of which can play a role in disease processes. Alteration of their activity is a means to alter the disease phenotype and as such identification of novel adhesion molecules is highly relevant as they may play a role in many diseases, particularly inflammatory disease, oncology, and cardiovascular disease.

Table 1. Integrins:

<i>Integrin Receptor</i>	<i>Ligand</i>	<i>Distribution</i>
$\beta 1$ (CD29)		
$\alpha 1\beta 1$	Laminin, Collagen	Activated T cells, fibroblasts
$\alpha 2\beta 1$	Collagen, Laminin	Activated T cells, endothelial cells, platelets, basophils.
$\alpha 3\beta 1$	Laminin, Collagen, Basement membrane Fibronectin	
$\alpha 4\beta 1$	VCAM-1 (domains 1 and 4), Fibronectin domain), MadCAM-1	Lymphocytes, monocytes, eosinophils, (CS-1 basophils, mast cells, NK cells
$\alpha 5\beta 1$	Fibronectin	Lymphocytes, monocytes, endothelial cells, basophils, mast cells, fibroblasts
$\alpha 6\beta 1$	Laminin	Platelets, T cells, eosinophils, monocytes,

<i>Integrin Receptor</i>	<i>Ligand</i>	<i>Distribution</i>
		endothelial cells
$\alpha 9\beta 1$	Tenascin, Osteopontin	VCAM-1, Airway epithelial cells, smooth muscle cells, neutrophils
$\alpha V\beta 1$	Vitronectin, fibronectin	Platelets, B cells.
$\beta 2$ (CD18)		
LFA-1	ICAM-1, 2, 3	All leukocytes
(CD11a/CD18)		
Mac-1	ICAM-1, Fibrinogen, LPS	Granulocytes, monocytes
(CD11b/CD18		
)		
αD	ICAM-3, VCAM-1	Tissue macrophages, monocytes, CD8+ T cells, eosinophils
$\beta 3$ (CD61)		
GpIIb/IIIa	Fibrinogen, Vitronectin, vWF	Platelets, endothelial cells
$\alpha V/IIIa$	Vitronectin, Fibrinogen, vWF, Laminin, Thrombospondin, Osteopontin	Platelets,
$\beta 7$		
$\alpha 4\beta 7$	MAdCAM-1, VCAM-1	Subset of memory T cells, eosinophils,
(LPAM-1)	Firbonectin (CS-1 domain)	basophils, endothelial cells
$\alpha E\beta 7$	E-cadherin	Intestinal intraepithelial lymphocytes.

Table 2. Selectins:

Receptor	Ligand	Distribution
E-selectin	Sialyl-LewisX, selectin, LFA-1, ESL-1, PSGL-1	L- Activated endothelial cells
L-selectin	GlyCAM-1, MAdCAM-1, CD34, Sialyl LewisX, E-selectin, P-selectin	Resting leukocytes
P-selectin	Sialyl-LewisX, selectin, PSGL-1	L- Activated endothelial cells, activated platelets

Table 3. Immunoglobulin superfamily:

Receptor	Ligand	Distribution
ICAM-1 5 Ig domains	LFA-1 (CD11a/CD18) Mac-1 (CD11b/CD18), CD43	Widespread, endothelial cells, fibroblasts, epithelium, monocytes, lymphocytes, dendritic cells, chondrocytes.
ICAM-2 2 Ig domains	LFA-1 (CD11b)	endothelial cells (high): lymphocytes, monocytes, basophils, platelets (low).
ICAM-3 5 Ig domains	LFA-1 (α d/CD18)	Lymphocytes, monocytes, neutrophils, eosinophils, basophils.
VCAM-1 6 or 7 Ig domains	α 4 β 1, α 4 β 7	Endothelial cells, monocytes, fibroblasts, dendritic cells, bone marrow stromal cells, myoblasts.
LFA-3 6 Ig domains	CD2	Endothelial cells, leukocytes, epithelial cells
PECAM-1	CD31, heparin	Endothelial cells (at EC-EC junctions), T cell

(CD31)		subsets, platelets, neutrophils, eosinophils, monocytes, smooth muscle cells, bone marrow stem cells.
NCAM	NCAM, heparin SO ₄	Neural cells, muscle
MadCAM-1	$\alpha 4\beta 7$, L-selectin	Peyer's patch, mesenteric lymph nodes, mucosal endothelial cells, spleen.
4 Ig domains		
CD2	CD58, CD59, CD48	T lymphocytes

THE INVENTION

The invention is based on the discovery that the AAC74854.1 protein, AAC76768.1 protein, AAD05645.1 protein, VP4 protein and Ro60 protein function as adhesion molecules.

In a first aspect, the invention provides a polypeptide, which polypeptide:

- (i) has the amino acid sequence as recited in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10;
- (ii) is a fragment thereof having adhesion molecule activity or having an antigenic determinant in common with the polypeptides of (i); or
- (iii) is a functional equivalent of (i) or (ii).

The polypeptide having the sequence recited in SEQ ID NO: 2 is referred to hereafter as "the ADS1 polypeptide". The polypeptide having the sequence recited in SEQ ID NO: 4 is referred to hereafter as "the ADS2 polypeptide". The polypeptide having the sequence recited in SEQ ID NO: 6 is referred to hereafter as "the ADS3 polypeptide". The polypeptide having the sequence recited in SEQ ID NO: 8 is referred to hereafter as "the ADS4 polypeptide". The polypeptide having the sequence recited in SEQ ID NO: 10 is referred to hereafter as "the ADS5 polypeptide".

In a second aspect, the invention provides a purified nucleic acid molecule, which encodes a polypeptide of the first aspect of the invention. Preferably, the purified nucleic

acid molecule has the nucleic acid sequence as recited in SEQ ID NO: 1 (encoding the ADS1 polypeptide), SEQ ID NO: 3 (encoding the ADS2 polypeptide), SEQ ID NO: 5 (encoding the ADS3 polypeptide) SEQ ID NO: 7 (encoding the ADS4 polypeptide), SEQ ID NO: 9 (encoding the ADS5 polypeptide) or is a redundant equivalent or fragment of either of these sequences.

In a third aspect, the invention provides a purified nucleic acid molecule, which hybridizes under high stringency conditions with a nucleic acid molecule of the second aspect of the invention.

In a fourth aspect, the invention provides a vector, such as an expression vector, that contains a nucleic acid molecule of the second or third aspect of the invention.

In a fifth aspect, the invention provides a host cell transformed with a vector of the fourth aspect of the invention.

In a sixth aspect, the invention provides a ligand which binds specifically to, and which preferably inhibits the adhesion molecule activity of, a polypeptide of the first aspect of the invention.

In a seventh aspect, the invention provides a compound that is effective to alter the expression of a natural gene which encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

A compound of the seventh aspect of the invention may either increase (agonise) or decrease (antagonise) the level of expression of the gene or the activity of the polypeptide. Importantly, the identification of the function of the ADS1, ADS2, ADS3, ADS4 and ADS5 polypeptides allows for the design of screening methods capable of identifying compounds that are effective in the treatment and/or diagnosis of disease.

In an eighth aspect, the invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the fifth aspect of the invention, or a compound of the sixth aspect of the invention, for use in therapy or diagnosis. These molecules may also be used in the manufacture of a medicament for the treatment of cardiovascular disease, cancer, asthma, COPD, inflammatory disease or bacterial

infections.

In a ninth aspect, the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide of the first aspect of the invention or the activity of a polypeptide of the first aspect of the invention in tissue from said patient and comparing said level of expression or activity to
5 a control level, wherein a level that is different to said control level is indicative of disease. Such a method will preferably be carried out in vitro. Similar methods may be used for monitoring the therapeutic treatment of disease in a patient, wherein altering the level of expression or activity of a polypeptide or nucleic acid molecule over the period
10 of time towards a control level is indicative of regression of disease.

The adhesion molecules whose sequences are presented in SEQ ID NO: 2 and SEQ ID NO: 4 are implicated herein in the pathogenicity of the organism *Escherichia Coli*. Accordingly, ligands to these proteins are likely to be effective in controlling disease caused by this organism. Furthermore, these proteins provide a potential component for a
15 vaccine against this organism and the diseases that it causes.

The adhesion molecule whose sequence is presented in SEQ ID NO: 6 is implicated herein in the pathogenicity of the organism *Helicobacter Pylori*. Accordingly, ligands to this protein are likely to be effective in controlling disease caused by this organism. Furthermore, this protein provides a potential component for a vaccine against this
20 organism and the diseases that it causes.

The adhesion molecule whose sequence is presented in SEQ ID NO: 8 is implicated herein in the pathogenicity of the organism *Human Rotavirus*. Accordingly, ligands to this protein are likely to be effective in controlling disease caused by this organism. Furthermore, this protein provides a potential component for a vaccine against this
25 organism and the diseases that it causes.

A preferred method for detecting polypeptides of the first aspect of the invention comprises the steps of: (a) contacting a ligand, such as an antibody, of the sixth aspect of the invention with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

A number of different such methods according to the ninth aspect of the invention exist, as the skilled reader will be aware, such as methods of nucleic acid hybridization with short probes, point mutation analysis, polymerase chain reaction (PCR) amplification and methods using antibodies to detect aberrant protein levels. Similar methods may be used on a short or long-term basis to allow therapeutic treatment of a disease to be monitored in a patient. The invention also provides kits that are useful in these methods for diagnosing disease.

In a tenth aspect, the invention provides for the use of a polypeptide of the first aspect of the invention as an adhesion molecule.

In an eleventh aspect, the invention provides a pharmaceutical composition comprising a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, in conjunction with a pharmaceutically-acceptable carrier.

In a twelfth aspect, the present invention provides a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention, for use in the manufacture of a medicament for the diagnosis or treatment of a disease, such as herpes virus infection.

In a thirteenth aspect, the invention provides a method of treating a disease in a patient comprising administering to the patient a polypeptide of the first aspect of the invention, or a nucleic acid molecule of the second or third aspect of the invention, or a vector of the fourth aspect of the invention, or a ligand of the sixth aspect of the invention, or a compound of the seventh aspect of the invention.

For diseases in which the expression of a natural gene encoding a polypeptide of the first aspect of the invention, or in which the activity of a polypeptide of the first aspect of the invention, is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an agonist. Conversely, for diseases in which the

expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, ligand or compound administered to the patient should be an antagonist. Examples of such antagonists include antisense nucleic acid molecules, ribozymes and ligands, such as antibodies.

5 In a fourteenth aspect, the invention provides transgenic or knockout non-human animals that have been transformed to express higher, lower or absent levels of a polypeptide of the first aspect of the invention. Such transgenic animals are very useful models for the study of disease and may also be using in screening regimes for the identification of compounds that are effective in the treatment or diagnosis of such a disease.

10 A summary of standard techniques and procedures, which may be employed in order to utilise the invention, is given below. It will be understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors and reagents described. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only and it is not intended that this terminology should limit the scope of the present invention. The extent of the invention is limited only by the terms of the appended claims.

Standard abbreviations for nucleotides and amino acids are used in this specification.

20 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA technology and immunology, which are within the skill of those working in the art.

Such techniques are explained fully in the literature. Examples of particularly suitable texts for consultation include the following: Sambrook Molecular Cloning; A Laboratory Manual, Second Edition (1989); DNA Cloning, Volumes I and II (D.N Glover ed. 1985);
25 Oligonucleotide Synthesis (M.J. Gait ed. 1984); Nucleic Acid Hybridization (B.D. Hames & S.J. Higgins eds. 1984); Transcription and Translation (B.D. Hames & S.J. Higgins eds. 1984); Animal Cell Culture (R.I. Freshney ed. 1986); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide to Molecular Cloning (1984);
the Methods in Enzymology series (Academic Press, Inc.), especially volumes 154 &
30 155; Gene Transfer Vectors for Mammalian Cells (J.H. Miller and M.P. Calos eds. 1987,

Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds. 1987, Academic Press, London); Scopes, (1987) Protein Purification: Principles and Practice, Second Edition (Springer Verlag, N.Y.); and Handbook of Experimental Immunology, Volumes I-IV (D.M. Weir and C. C. Blackwell eds. 1986).

As used herein, the term "polypeptide" includes any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e. peptide isosteres. This term refers both to short chains (peptides and oligopeptides) and to longer chains (proteins).

10 The polypeptide of the present invention may be in the form of a mature protein or may be a pre-, pro- or prepro- protein that can be activated by cleavage of the pre-, pro- or prepro- portion to produce an active mature polypeptide. In such polypeptides, the pre-, pro- or prepro- sequence may be a leader or secretory sequence or may be a sequence that is employed for purification of the mature polypeptide sequence.

15 The polypeptide of the first aspect of the invention may form part of a fusion protein. For example, it is often advantageous to include one or more additional amino acid sequences which may contain secretory or leader sequences, pro-sequences, sequences which aid in purification, or sequences that confer higher protein stability, for example during recombinant production. Alternatively or additionally, the mature polypeptide may be
20 fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol).

Polypeptides may contain amino acids other than the 20 gene-encoded amino acids, modified either by natural processes, such as by post-translational processing or by chemical modification techniques which are well known in the art. Among the known
25 modifications which may commonly be present in polypeptides of the present invention are glycosylation, lipid attachment, sulphation, gamma-carboxylation, for instance of glutamic acid residues, hydroxylation and ADSP-ribosylation. Other potential modifications include acetylation, acylation, amidation, covalent attachment of flavin, covalent attachment of a haeme moiety, covalent attachment of a nucleotide or nucleotide
30 derivative, covalent attachment of a lipid derivative, covalent attachment of

phosphatidylinositol, cross-linking, cyclization, disulphide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, GPI anchor formation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl terminus in a polypeptide, or both, by a covalent modification is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention.

The modifications that occur in a polypeptide often will be a function of how the polypeptide is made. For polypeptides that are made recombinantly, the nature and extent of the modifications in large part will be determined by the post-translational modification capacity of the particular host cell and the modification signals that are present in the amino acid sequence of the polypeptide in question. For instance, glycosylation patterns vary between different types of host cell.

The polypeptides of the present invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides (for example purified from cell culture), recombinantly-produced polypeptides (including fusion proteins), synthetically-produced polypeptides or polypeptides that are produced by a combination of these methods.

The functionally equivalent polypeptides of the first aspect of the invention may be polypeptides that are homologous to the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptides. Two polypeptides are said to be "homologous", as the term is used herein, if the sequence of one of the polypeptides has a high enough degree of identity or similarity to the sequence of the other polypeptide. "Identity" indicates that at any particular position in the aligned sequences, the amino acid residue is identical between the sequences. "Similarity" indicates that, at any particular position in the aligned sequences, the amino acid residue is of a similar type between the sequences. Degrees of

etc. Monoclonal antibodies are particularly useful for identifying polypeptides against which they are directed. Monoclonal antibodies of interest may be isolated using techniques known in the art, and cloned and expressed.

- 5 Chimeric antibodies, in which non-human variable regions and constant regions (see, for example, Liu et al., *Proc. Natl. Acad. Sci. USA*, 84, 3323 (1987)), may also be of use.

The antibody may be modified to make it less immunogenic by humanisation (see Jones et al., *Nature*, 321, 529-532 (1986); 239, 1534 (1988); Kabat et al., *J. Immunol.*, 147, 1302-1306 (1991); Acad. Sci. USA, 86, 10029 (1989); Gorman et al., *Proc. Natl. Acad. Sci. USA*, 88, 1320-1324 (1991); and Hodgson et al., *Bio/Technology*, 9, 43-46 (1991)). The term "humanised antibody", as used herein, refers to antibody molecules in which one or more amino acids in the variable domain and selected other amino acids in the variable domain of a non-human donor antibody have been substituted with corresponding amino acids in a human antibody. The humanised antibody thus retains the binding ability of the donor antibody but has the binding ability of the donor antibody.

In a further alternative, the antibody may be a bispecific antibody having two different antigen-binding domains, each specific for a different epitope.

Phage display technology may be utilised to select antibodies with specific binding activities towards the polypeptides of the target. PCR amplified V-genes of lymphocytes from immunised animals, or from naive libraries (McCafferty et al., *Nature*, 348, 552-554; Marks, J. et al., (1992) *Biotechnology*, 10, 779-783; antibodies can also be improved by chain shuffling (see, for example, Marks et al., 352, 624-628).

Antibodies generated by the above techniques, with or without additional utility in that they may be employed in immunoassays, radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA).

THIS PAGE BLANK (USPTO)

identity and similarity can be readily calculated (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing. Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991).

Homologous polypeptides therefore include natural biological variants (for example, allelic variants or geographical variations within the species from which the polypeptides are derived) and mutants (such as mutants containing amino acid substitutions, insertions or deletions) of the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptides. Such mutants may include polypeptides in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; among the basic residues Lys and Arg; or among the aromatic residues Phe and Tyr. Particularly preferred are variants in which several, i.e. between 5 and 10, 1 and 5, 1 and 3, 1 and 2 or just 1 amino acids are substituted, deleted or added in any combination. Especially preferred are silent substitutions, additions and deletions, which do not alter the properties and activities of the protein. Also especially preferred in this regard are conservative substitutions.

Such mutants also include polypeptides in which one or more of the amino acid residues include a substituent group;

Typically, greater than 50% identity between two polypeptides is considered to be an indication of functional equivalence. Preferably, functionally equivalent polypeptides of the first aspect of the invention have a degree of sequence identity with the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptide or with active fragments thereof, of greater than 50%. More preferred polypeptides have degrees of identity of greater than 60%, 70%, 80%, 90%, 95%, 98% or 99%, respectively.

The functionally equivalent polypeptides of the first aspect of the invention may also be polypeptides, which have been identified using one or more techniques of structural alignment. For example, the Inpharmatica Genome Threader technology that forms one aspect of the search tools used to generate the Biopendium search database may be used
5 (see co-pending United Kingdom patent application GB0006153.1) to identify polypeptides of presently-unknown function which, while having low sequence identity as compared to the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptides, are predicted to have adhesion molecule activity, by virtue of sharing significant structural homology with the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptide sequences. By "significant
10 structural homology" is meant that the Inpharmatica Genome Threader predicts two proteins to share structural homology with a certainty of 80% and above.

The polypeptides of the first aspect of the invention also include fragments of the ADS1, ADS2, ADS3, ADS4 and ADS5 polypeptides and fragments of the functional equivalents of the ADS1, ADS2, ADS3, ADS4 and ADS5 polypeptides, provided that those
15 fragments retain adhesion molecule activity or have an antigenic determinant in common with the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptides.

As used herein, the term "fragment" refers to a polypeptide having an amino acid sequence that is the same as part, but not all, of the amino acid sequence of the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptides or one of its functional equivalents. The
20 fragments should comprise at least n consecutive amino acids from the sequence and, depending on the particular sequence, n preferably is 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more). Small fragments may form an antigenic determinant.

Such fragments may be "free-standing", i.e. not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a
25 part or region. When comprised within a larger polypeptide, the fragment of the invention most preferably forms a single continuous region. For instance, certain preferred embodiments relate to a fragment having a pre - and/or pro- polypeptide region fused to the amino terminus of the fragment and/or an additional region fused to the carboxyl terminus of the fragment. However, several fragments may be comprised within a single
30 larger polypeptide.

The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones expressing the polypeptides of the invention or to purify the polypeptides by affinity chromatography. The antibodies may also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')₂ and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the first aspect of the invention.

If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The polypeptide used to immunise the animal can be derived by recombinant DNA technology or can be synthesized chemically. If desired, the polypeptide can be conjugated to a carrier protein. Commonly used carriers to which the polypeptides may be chemically coupled include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor et al., *Immunology Today* 4: 72 (1983); Cole et al., 77-96 in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985).

Panels of monoclonal antibodies produced against the polypeptides of the first aspect of the invention can be screened for various properties, i.e., for isotype, epitope, affinity,

The polypeptides of the present invention or their immunogenic fragments (comprising at least one antigenic determinant) can be used to generate ligands, such as polyclonal or monoclonal antibodies, that are immunospecific for the polypeptides. Such antibodies may be employed to isolate or to identify clones expressing the polypeptides of the invention or to purify the polypeptides by affinity chromatography. The antibodies may also be employed as diagnostic or therapeutic aids, amongst other applications, as will be apparent to the skilled reader.

The term "immunospecific" means that the antibodies have substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art. As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')₂ and Fv, which are capable of binding to the antigenic determinant in question. Such antibodies thus bind to the polypeptides of the first aspect of the invention.

If polyclonal antibodies are desired, a selected mammal, such as a mouse, rabbit, goat or horse, may be immunised with a polypeptide of the first aspect of the invention. The polypeptide used to immunise the animal can be derived by recombinant DNA technology or can be synthesized chemically. If desired, the polypeptide can be conjugated to a carrier protein. Commonly used carriers to which the polypeptides may be chemically coupled include bovine serum albumin, thyroglobulin and keyhole limpet haemocyanin. The coupled polypeptide is then used to immunise the animal. Serum from the immunised animal is collected and treated according to known procedures, for example by immunoaffinity chromatography.

Monoclonal antibodies to the polypeptides of the first aspect of the invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies using hybridoma technology is well known (see, for example, Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor et al., *Immunology Today* 4: 72 (1983); Cole et al., 77-96 in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc. (1985).

Panels of monoclonal antibodies produced against the polypeptides of the first aspect of the invention can be screened for various properties, i.e., for isotype, epitope, affinity,

etc. Monoclonal antibodies are particularly useful in purification of the individual polypeptides against which they are directed. Alternatively, genes encoding the monoclonal antibodies of interest may be isolated from hybridomas, for instance by PCR techniques known in the art, and cloned and expressed in appropriate vectors.

- 5 Chimeric antibodies, in which non-human variable regions are joined or fused to human constant regions (see, for example, Liu et al., Proc. Natl. Acad. Sci. USA, 84, 3439 (1987)), may also be of use.

The antibody may be modified to make it less immunogenic in an individual, for example by humanisation (see Jones et al., Nature, 321, 522 (1986); Verhoeven et al., Science, 10 239, 1534 (1988); Kabat et al., J. Immunol., 147, 1709 (1991); Queen et al., Proc. Natl Acad. Sci. USA, 86, 10029 (1989); Gorman et al., Proc. Natl Acad. Sci. USA, 88, 34181 (1991); and Hodgson et al., Bio/Technology, 9, 421 (1991)). The term "humanised antibody", as used herein, refers to antibody molecules in which the CDR amino acids and selected other amino acids in the variable domains of the heavy and/or light chains of 15 a non-human donor antibody have been substituted in place of the equivalent amino acids in a human antibody. The humanised antibody thus closely resembles a human antibody but has the binding ability of the donor antibody.

In a further alternative, the antibody may be a "bispecific" antibody, that is an antibody having two different antigen-binding domains, each domain being directed against a 20 different epitope.

Phage display technology may be utilised to select genes which encode antibodies with binding activities towards the polypeptides of the invention either from repertoires of PCR amplified V-genes of lymphocytes from humans screened for possessing the relevant antibodies, or from naive libraries (McCafferty, J. et al., (1990), Nature 348, 25 552-554; Marks, J. et al., (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) Nature 352, 624-628).

Antibodies generated by the above techniques, whether polyclonal or monoclonal, have additional utility in that they may be employed as reagents in immunoassays, 30 radioimmunoassays (RIA) or enzyme-linked immunosorbent assays (ELISA). In these

applications, the antibodies can be labelled with an analytically detectable reagent such as a radioisotope, a fluorescent molecule or an enzyme.

Preferred nucleic acid molecules of the second and third aspects of the invention are those which encode the polypeptide sequences recited in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 and SEQ ID NO: 10 , and functionally equivalent polypeptides. These nucleic acid molecules may be used in the methods and applications described herein. The nucleic acid molecules of the invention preferably comprise at least n consecutive nucleotides from the sequences disclosed herein where, depending on the particular sequence, n is 10 or more (for example, 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

The nucleic acid molecules of the invention also include sequences that are complementary to nucleic acid molecules described above (for example, for antisense or probing purposes).

Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance cDNA, synthetic DNA or genomic DNA. Such nucleic acid molecules may be obtained by cloning, by chemical synthetic techniques or by a combination thereof. The nucleic acid molecules can be prepared, for example, by chemical synthesis using techniques such as solid phase phosphoramidite chemical synthesis, from genomic or cDNA libraries or by separation from an organism. RNA molecules may generally be generated by the in vitro or in vivo transcription of DNA sequences.

The nucleic acid molecules may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The term "nucleic acid molecule" also includes analogues of DNA and RNA, such as those containing modified backbones, and peptide nucleic acids (PNA). The term "PNA", as used herein refers to an antisense molecule or an anti-gene agent that comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues, which preferably ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in a cell,

where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8: 53-63).

A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 2 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 1. A
5 nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 4 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 3. . A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 6 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 5. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 8 may be identical to the coding
10 sequence of the nucleic acid molecule shown in SEQ ID NO: 7. A nucleic acid molecule which encodes the polypeptide of SEQ ID NO: 10 may be identical to the coding sequence of the nucleic acid molecule shown in SEQ ID NO: 9.

These molecules also may have a different sequence which, as a result of the degeneracy of the genetic code, encodes a polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID
15 NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10. Such nucleic acid molecules that encode the polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10 may include, but are not limited to, the coding sequence for the mature polypeptide by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pro-,
20 pre- or prepro- polypeptide sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with further additional, non-coding sequences, including non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals), ribosome binding and mRNA stability. The nucleic acid molecules
25 may also include additional sequences, which encode additional amino acids, such as those, which provide additional functionalities.

The nucleic acid molecules of the second and third aspects of the invention may also encode the fragments or the functional equivalents of the polypeptides and fragments of the first aspect of the invention. Such a nucleic acid molecule may be a naturally
30 occurring variant such as a naturally occurring allelic variant, or the molecule may be a

variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells or organisms.

5 Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or insertions. The substitutions, deletions or insertions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or insertions.

10 The nucleic acid molecules of the invention can also be engineered, using methods generally known in the art, for a variety of reasons, including modifying the cloning, processing, and/or expression of the gene product (the polypeptide). DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides are included as techniques, which may be used to engineer the nucleotide sequences. Site-directed mutagenesis may be used to insert new restriction
15 sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations and so forth.

Nucleic acid molecules, which encode a polypeptide of the first aspect of the invention, may be ligated to a heterologous sequence so that the combined nucleic acid molecule encodes a fusion protein. Such combined nucleic acid molecules are included within the
20 second or third aspects of the invention. For example, to screen peptide libraries for inhibitors of the activity of the polypeptide, it may be useful to express, using such a combined nucleic acid molecule, a fusion protein that can be recognised by a commercially-available antibody. A fusion protein may also be engineered to contain a cleavage site located between the sequence of the polypeptide of the invention and the
25 sequence of a heterologous protein so that the polypeptide may be cleaved and purified away from the heterologous protein.

The nucleic acid molecules of the invention also include antisense molecules that are partially complementary to nucleic acid molecules encoding polypeptides of the present invention and that therefore hybridize to the encoding nucleic acid molecules
30 (hybridization). Such antisense molecules, such as oligonucleotides, can be designed to

recognise, specifically bind to and prevent transcription of a target nucleic acid encoding a polypeptide of the invention, as will be known by those of ordinary skill in the art (see, for example, Cohen, J.S., Trends in Pharm. Sci., 10, 435 (1989), Okano, J. Neurochem. 56, 560 (1991); O'Connor, J. Neurochem 56, 560 (1991); Lee et al., Nucleic Acids Res 6, 3073 (1979); Cooney et al., Science 241, 456 (1988); Dervan et al., Science 251, 1360 (1991).

The term "hybridization" as used here refers to the association of two nucleic acid molecules with one another by hydrogen bonding. Typically, one molecule will be fixed to a solid support and the other will be free in solution. Then, the two molecules may be placed in contact with one another under conditions that favour hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase molecule to the solid support (Denhardt's reagent or BLOTTO); the concentration of the molecules; use of compounds to increase the rate of association of molecules (dextran sulphate or polyethylene glycol); and the stringency of the washing conditions following hybridization (see Sambrook et al. [supra]).

The inhibition of hybridization of a completely complementary molecule to a target molecule may be examined using a hybridization assay, as known in the art (see, for example, Sambrook et al [supra]). A substantially homologous molecule will then compete for and inhibit the binding of a completely homologous molecule to the target molecule under various conditions of stringency, as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152: 399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152: 507-511).

"Stringency" refers to conditions in a hybridization reaction that favour the association of very similar molecules over association of molecules that differ. High stringency hybridisation conditions are defined as overnight incubation at 42°C in a solution comprising 50% formamide, 5XSSC (150mM NaCl, 15mM trisodium citrate), 50mM sodium phosphate (pH7.6), 5x Denhardts solution, 10% dextran sulphate, and 20 microgram/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1X SSC at approximately 65°C. Low stringency conditions involve the hybridisation

reaction being carried out at 35°C (see Sambrook et al. [supra]). Preferably, the conditions used for hybridization are those of high stringency.

Preferred embodiments of this aspect of the invention are nucleic acid molecules that are at least 70% identical over their entire length to a nucleic acid molecule encoding the ADS1 polypeptide (SEQ ID NO: 2), ADS2 polypeptide (SEQ ID NO: 4), ADS3 polypeptide (SEQ ID NO: 6), ADS4 polypeptide (SEQ ID NO: 8), ADS5 polypeptide (SEQ ID NO: 10) and nucleic acid molecules that are substantially complementary to such nucleic acid molecules. Preferably, a nucleic acid molecule according to this aspect of the invention comprises a region that is at least 80% identical over its entire length to the nucleic acid molecule having the sequence given in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 or a nucleic acid molecule that is complementary thereto. In this regard, nucleic acid molecules at least 90%, preferably at least 95%, more preferably at least 98% or 99% identical over their entire length to the same are particularly preferred. Preferred embodiments in this respect are nucleic acid molecules that encode polypeptides that retain substantially the same biological function or activity as the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptides.

The invention also provides a process for detecting a nucleic acid molecule of the invention, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting any such duplexes that are formed.

As discussed additionally below in connection with assays that may be utilised according to the invention, a nucleic acid molecule as described above may be used as a hybridization probe for RNA, cDNA or genomic DNA, in order to isolate full-length cDNAs and genomic clones encoding the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptides and to isolate cDNA and genomic clones of homologous or orthologous genes that have a high sequence similarity to the gene encoding this polypeptide.

In this regard, the following techniques, among others known in the art, may be utilised and are discussed below for purposes of illustration. Methods for DNA sequencing and analysis are well known and are generally available in the art and may, indeed, be used to practice many of the embodiments of the invention discussed herein. Such methods may

employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proof-reading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the sequencing process may be automated using machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), the Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

One method for isolating a nucleic acid molecule encoding a polypeptide with an equivalent function to that of the ADS1, ADS2, ADS3, ADS4 or ADS5 polypeptides is to probe a genomic or cDNA library with a natural or artificially-designed probe using standard procedures that are recognised in the art (see, for example, "Current Protocols in Molecular Biology", Ausubel et al. (eds). Greene Publishing Association and John Wiley Interscience, New York, 1989,1992). Probes comprising at least 15, preferably at least 30, and more preferably at least 50, contiguous bases that correspond to, or are complementary to, nucleic acid sequences from the appropriate encoding gene (SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 or SEQ ID NO: 9), are particularly useful probes. Such probes may be labelled with an analytically detectable reagent to facilitate their identification. Useful reagents include, but are not limited to, radioisotopes, fluorescent dyes and enzymes that are capable of catalysing the formation of a detectable product. Using these probes, the ordinarily skilled artisan will be capable of isolating complementary copies of genomic DNA, cDNA or RNA polynucleotides encoding proteins of interest from human, mammalian or other animal sources and screening such sources for related sequences, for example, for additional members of the family, type and/or subtype.

In many cases, isolated cDNA sequences will be incomplete, in that the region encoding the polypeptide will be cut short, normally at the 5' end. Several methods are available to obtain full length cDNAs, or to extend short cDNAs. Such sequences may be extended utilising a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed is based on the method of Rapid Amplification of cDNA

Ends (RACE; see, for example, Frohman et al., PNAS USA 85, 8998-9002, 1988). Recent modifications of this technique, exemplified by the MarathonTM technology (Clontech Laboratories Inc.), for example, have significantly simplified the search for longer cDNAs. A slightly different technique, termed "restriction-site" PCR, uses universal primers to retrieve unknown nucleic acid sequence adjacent a known locus (Sarkar, G. (1993) PCR Methods Applic. 2: 318-322). Inverse PCR may also be used to amplify or to extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16: 8186). Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic., 1, 111-119). Another method which may be used to retrieve unknown sequences is that of Parker, J.D. et al. (1991); Nucleic Acids Res. 19: 3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinderTM libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences that contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

In one embodiment of the invention, the nucleic acid molecules of the present invention may be used for chromosome localisation. In this technique, a nucleic acid molecule is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important step in the confirmatory correlation of those sequences with the gene-associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationships between genes and diseases that

have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes). This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localised by genetic linkage to a particular genomic region, any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleic acid molecule may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

The nucleic acid molecules of the present invention are also valuable for tissue localisation. Such techniques allow the determination of expression patterns of the polypeptide in tissues by detection of the mRNAs that encode them. These techniques include in situ hybridization techniques and nucleotide amplification techniques, such as PCR. Results from these studies provide an indication of the normal functions of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by a mutant gene provide valuable insights into the role of mutant polypeptides in disease. Such inappropriate expression may be of a temporal, spatial or quantitative nature.

The vectors of the present invention comprise nucleic acid molecules of the invention and may be cloning or expression vectors. The host cells of the invention, which may be transformed, transfected or transduced with the vectors of the invention may be prokaryotic or eukaryotic.

The polypeptides of the invention may be prepared in recombinant form by expression of their encoding nucleic acid molecules in vectors contained within a host cell. Such expression methods are well known to those of skill in the art and many are described in detail by Sambrook et al (supra) and Fernandez & Hoeffler (1998, eds. "Gene expression systems. Using nature for the art of expression". Academic Press, San Diego, London, Boston, New York, Sydney, Tokyo, Toronto).

Generally, any system or vector that is suitable to maintain, propagate or express nucleic acid molecules to produce a polypeptide in the required host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a

variety of well-known and routine techniques, such as, for example, those described in Sambrook et al., (supra). Generally, the encoding gene can be placed under the control of a control element such as a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator, so that the DNA sequence encoding the desired polypeptide is transcribed into RNA in the transformed host cell.

Examples of suitable expression systems include, for example, chromosomal, episomal and virus-derived systems, including, for example, vectors derived from: bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, or combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, including cosmids and phagemids. Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid.

Particularly suitable expression systems include microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (for example, baculovirus); plant cell systems transformed with virus expression vectors (for example, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (for example, Ti or pBR322 plasmids); or animal cell systems. Cell-free translation systems can also be employed to produce the polypeptides of the invention.

Introduction of nucleic acid molecules encoding a polypeptide of the present invention into host cells can be effected by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook et al., [supra]. Particularly suitable methods include calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection (see Sambrook et al., 1989 [supra]; Ausubel et al., 1991 [supra]; Spector, Goldman & Leinwald, 1998). In eukaryotic cells, expression systems may either be

transient (for example, episomal) or permanent (chromosomal integration) according to the needs of the system.

The encoding nucleic acid molecule may or may not include a sequence encoding a control sequence, such as a signal peptide or leader sequence, as desired, for example, for secretion of the translated polypeptide into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals. Leader sequences can be removed by the bacterial host in post-translational processing.

In addition to control sequences, it may be desirable to add regulatory sequences that allow for regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory sequences are those which cause the expression of a gene to be increased or decreased in response to a chemical or physical stimulus, including the presence of a regulatory compound or to various temperature or metabolic conditions. Regulatory sequences are those non-translated regions of the vector, such as enhancers, promoters and 5' and 3' untranslated regions. These interact with host cellular proteins to carry out transcription and translation. Such regulatory sequences may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript phagemid (Stratagene, LaJolla, CA) or pSport1™ plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (for example, heat shock, RUBISCO and storage protein genes) or from plant viruses (for example, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

An expression vector is constructed so that the particular nucleic acid coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and

orientation of the coding sequence with respect to the regulatory sequences being such that the coding sequence is transcribed under the "control" of the regulatory sequences, i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence. In some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame.

The control sequences and other regulatory sequences may be ligated to the nucleic acid coding sequence prior to insertion into a vector. Alternatively, the coding sequence can be cloned directly into an expression vector that already contains the control sequences and an appropriate restriction site.

For long-term, high-yield production of a recombinant polypeptide, stable expression is preferred. For example, cell lines, which stably express the polypeptide of interest, may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalised cell lines available from the American Type Culture Collection (ATCC) including, but not limited to, Chinese hamster ovary (CHO), HeLa, baby hamster kidney (BHK), monkey kidney (COS), C127, 3T3, BHK, HEK 293, Bowes melanoma and human hepatocellular carcinoma (for example Hep G2) cells and a number of other cell lines.

In the baculovirus system, the materials for baculovirus/insect cell expression systems are commercially available in kit form from, inter alia, Invitrogen, San Diego CA (the "MaxBac" kit). These techniques are generally known to those skilled in the art and are described fully in Summers and Smith, Texas Agricultural Experiment Station Bulletin

No. 1555 (1987). Particularly suitable host cells for use in this system include insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells.

There are many plant cell culture and whole plant genetic expression systems known in the art. Examples of suitable plant cellular genetic expression systems include those described in US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture have been described by Zenk, *Phytochemistry* 30, 3861-3863 (1991).

In particular, all plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be utilised, so that whole plants are recovered which contain the transferred gene. Practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugar cane, sugar beet, cotton, fruit and other trees, legumes and vegetables.

Examples of particularly preferred bacterial host cells include streptococci, staphylococci, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells.

Examples of particularly suitable host cells for fungal expression include yeast cells (for example, *S. cerevisiae*) and *Aspergillus* cells.

Any number of selection systems are known in the art that may be used to recover transformed cell lines. Examples include the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) *Cell* 11: 223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) *Cell* 22: 817-23) genes that can be employed in tk- or apt^r cells, respectively.

Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dihydrofolate reductase (DHFR) that confers resistance to methotrexate (Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77: 3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150: 1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described, examples of which will be clear to those of skill in the art.

Although the presence or absence of marker gene expression suggests that the gene of

interest is also present, its presence and expression may need to be confirmed. For example, if the relevant sequence is inserted within a marker gene sequence, transformed cells containing the appropriate sequences can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a polypeptide of the invention under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain a nucleic acid sequence encoding a polypeptide of the invention and which express said polypeptide may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassays, for example, fluorescence activated cell sorting (FACS) or immunoassay techniques (such as the enzyme-linked immunosorbent assay [ELISA] and radioimmunoassay [RIA]), that include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein (see Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. et al. (1983) J. Exp. Med, 158, 1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labelled hybridization or PCR probes for detecting sequences related to nucleic acid molecules encoding polypeptides of the present invention include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled polynucleotide. Alternatively, the sequences encoding the polypeptide of the invention may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesise RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH)).

Suitable reporter molecules or labels, which may be used for ease of detection, include

radionuclides, enzymes and fluorescent, chemiluminescent or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Nucleic acid molecules according to the present invention may also be used to create transgenic animals, particularly rodent animals. Such transgenic animals form a further aspect of the present invention. This may be done locally by modification of somatic cells, or by germ line therapy to incorporate heritable modifications. Such transgenic animals may be particularly useful in the generation of animal models for drug molecules effective as modulators of the polypeptides of the present invention.

The polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. High performance liquid chromatography is particularly useful for purification. Well-known techniques for refolding proteins may be employed to regenerate an active conformation when the polypeptide is denatured during isolation and or purification.

Specialised vector constructions may also be used to facilitate purification of proteins, as desired, by joining sequences encoding the polypeptides of the invention to a nucleotide sequence encoding a polypeptide domain that will facilitate purification of soluble proteins. Examples of such purification-facilitating domains include metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and the polypeptide of the invention may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the polypeptide of the invention fused to several histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification by IMAC (immobilised metal ion affinity chromatography as described in Porath, J. et al. (1992),

Prot. Exp. Purif. 3: 263-281) while the thioredoxin or enterokinase cleavage site provides a means for purifying the polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. et al. (1993; DNA Cell Biol. 12: 441-453).

- 5 If the polypeptide is to be expressed for use in screening assays, generally it is preferred that it be produced at the surface of the host cell in which it is expressed. In this event, the host cells may be harvested prior to use in the screening assay, for example using techniques such as fluorescence activated cell sorting (FACS) or immunoaffinity techniques. If the polypeptide is secreted into the medium, the medium can be recovered
- 10 in order to recover and purify the expressed polypeptide. If polypeptide is produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

The polypeptide of the invention can be used to screen libraries of compounds in any of a variety of drug screening techniques. Such compounds may activate (agonise) or inhibit (antagonise) the level of expression of the gene or the activity of the polypeptide of the invention and form a further aspect of the present invention. Preferred compounds are

15 effective to alter the expression of a natural gene that encodes a polypeptide of the first aspect of the invention or to regulate the activity of a polypeptide of the first aspect of the invention.

Agonist or antagonist compounds may be isolated from, for example, cells, cell-free preparations, chemical libraries or natural product mixtures. These agonists or antagonists

20 may be natural or modified substrates, ligands, enzymes, receptors or structural or functional mimetics. For a suitable review of such screening techniques, see Coligan et al., Current Protocols in Immunology 1(2): Chapter 5 (1991).

Compounds that are most likely to be good antagonists are molecules that bind to the polypeptide of the invention without inducing the biological effects of the polypeptide

25 upon binding to it. Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to the polypeptide of the invention and thereby inhibit or extinguish its activity. In this fashion, binding of the polypeptide to normal cellular binding molecules may be inhibited, such that the normal biological activity of

30 the polypeptide is prevented.

The polypeptide of the invention that is employed in such a screening technique may be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. In general, such screening procedures may involve using appropriate cells or cell membranes that express the polypeptide that are contacted with a test compound to observe binding, or stimulation or inhibition of a functional response. The functional response of the cells contacted with the test compound is then compared with control cells that were not contacted with the test compound. Such an assay may assess whether the test compound results in a signal generated by activation of the polypeptide, using an appropriate detection system. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist in the presence of the test compound is observed.

Alternatively, simple binding assays may be used, in which the adherence of a test compound to a surface bearing the polypeptide is detected by means of a label directly or indirectly associated with the test compound or in an assay involving competition with a labelled competitor. In another embodiment, competitive drug screening assays may be used, in which neutralising antibodies that are capable of binding the polypeptide specifically compete with a test compound for binding. In this manner, the antibodies can be used to detect the presence of any test compound that possesses specific binding affinity for the polypeptide.

Assays may also be designed to detect the effect of added test compounds on the production of mRNA encoding the polypeptide in cells. For example, an ELISA may be constructed that measures secreted or cell-associated levels of polypeptide using monoclonal or polyclonal antibodies by standard methods known in the art, and this can be used to search for compounds that may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues. The formation of binding complexes between the polypeptide and the compound being tested may then be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the polypeptide of interest (see International patent application WO84/03564). In this method, large numbers of

different small test compounds are synthesised on a solid substrate, which may then be reacted with the polypeptide of the invention and washed. One way of immobilising the polypeptide is to use non-neutralising antibodies. Bound polypeptide may then be detected using methods that are well known in the art. Purified polypeptide can also be coated directly onto plates for use in the aforementioned drug screening techniques.

The polypeptide of the invention may be used to identify membrane-bound or soluble receptors, through standard receptor binding techniques that are known in the art, such as ligand binding and crosslinking assays in which the polypeptide is labelled with a radioactive isotope, is chemically modified, or is fused to a peptide sequence that facilitates its detection or purification, and incubated with a source of the putative receptor (for example, a composition of cells, cell membranes, cell supernatants, tissue extracts, or bodily fluids). The efficacy of binding may be measured using biophysical techniques such as surface plasmon resonance and spectroscopy. Binding assays may be used for the purification and cloning of the receptor, but may also identify agonists and antagonists of the polypeptide, that compete with the binding of the polypeptide to its receptor. Standard methods for conducting screening assays are well understood in the art.

The invention also includes a screening kit useful in the methods for identifying agonists, antagonists, ligands, receptors, substrates, and enzymes, that are described above.

The invention includes the agonists, antagonists, ligands, receptors, substrates and enzymes, and other compounds which modulate the activity or antigenicity of the polypeptide of the invention discovered by the methods that are described above.

The invention also provides pharmaceutical compositions comprising a polypeptide, nucleic acid, ligand or compound of the invention in combination with a suitable pharmaceutical carrier. These compositions may be suitable as therapeutic or diagnostic reagents, as vaccines, or as other immunogenic compositions, as outlined in detail below.

According to the terminology used herein, a composition containing a polypeptide, nucleic acid, ligand or compound [X] is "substantially free of" impurities [herein, Y] when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more

preferably at least about 95%, 98% or even 99% by weight.

The pharmaceutical compositions should preferably comprise a therapeutically effective amount of the polypeptide, nucleic acid molecule, ligand, or compound of the invention. The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent needed to treat, ameliorate, or prevent a targetted disease or condition, or to exhibit a detectable therapeutic or preventative effect. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, for example, of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

The precise effective amount for a human subject will depend upon the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. This amount can be determined by routine experimentation and is within the judgement of the clinician. Generally, an effective dose will be from 0.01 mg/kg to 50 mg/kg, preferably 0.05 mg/kg to 10 mg/kg. Compositions may be administered individually to a patient or may be administered in combination with other agents, drugs or hormones.

A pharmaceutical composition may also contain a pharmaceutically acceptable carrier, for administration of a therapeutic agent. Such carriers include antibodies and other polypeptides, genes and other therapeutic agents such as liposomes, provided that the carrier does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolised macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers and inactive virus particles.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulphates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A

thorough discussion of pharmaceutically acceptable carriers is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may additionally contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such compositions. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

The pharmaceutical compositions utilised in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal or transcutaneous applications (for example, see WO98/20734), subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, intravaginal or rectal means. Gene guns or hyposprays may also be used to administer the pharmaceutical compositions of the invention. Typically, the therapeutic compositions may be prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared.

Direct delivery of the compositions will generally be accomplished by injection, subcutaneously, intraperitoneally, intravenously or intramuscularly, or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Dosage treatment may be a single dose schedule or a multiple dose schedule.

If the activity of the polypeptide of the invention is in excess in a particular disease state, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as described above, along with a pharmaceutically acceptable carrier in an amount effective to inhibit the function of the polypeptide, such as by blocking the binding of ligands, substrates, enzymes, receptors, or by inhibiting a second signal, and thereby alleviating the abnormal condition. Preferably, such

antagonists are antibodies. Most preferably, such antibodies are chimeric and/or humanised to minimise their immunogenicity, as described previously.

In another approach, soluble forms of the polypeptide that retain binding affinity for the ligand, substrate, enzyme, receptor, in question, may be administered. Typically, the polypeptide may be administered in the form of fragments that retain the relevant portions.

In an alternative approach, expression of the gene encoding the polypeptide can be inhibited using expression blocking techniques, such as the use of antisense nucleic acid molecules (as described above), either internally generated or separately administered.

Modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions (signal sequence, promoters, enhancers and introns) of the gene encoding the polypeptide. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Such oligonucleotides may be administered or may be generated in situ from expression in vivo.

In addition, expression of the polypeptide of the invention may be prevented by using ribozymes specific to its encoding mRNA sequence. Ribozymes are catalytically active RNAs that can be natural or synthetic (see for example Usman, N, et al., Curr. Opin. Struct. Biol (1996) 6(4), 527-33). Synthetic ribozymes can be designed to specifically cleave mRNAs at selected positions thereby preventing translation of the mRNAs into functional polypeptide. Ribozymes may be synthesised with a natural ribose phosphate backbone and natural bases, as normally found in RNA molecules. Alternatively the ribozymes may be synthesised with non-natural backbones, for example, 2'-O-methyl RNA, to provide protection from ribonuclease degradation and may contain modified

bases.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of non-traditional bases such as inosine, queosine and butosine, as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine and uridine which are not as easily recognised by endogenous endonucleases.

For treating abnormal conditions related to an under-expression of the polypeptide of the invention and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound that activates the polypeptide, i.e., an agonist as described above, to alleviate the abnormal condition. Alternatively, a therapeutic amount of the polypeptide in combination with a suitable pharmaceutical carrier may be administered to restore the relevant physiological balance of polypeptide.

Gene therapy may be employed to effect the endogenous production of the polypeptide by the relevant cells in the subject. Gene therapy is used to treat permanently the inappropriate production of the polypeptide by replacing a defective gene with a corrected therapeutic gene.

Gene therapy of the present invention can occur in vivo or ex vivo. Ex vivo gene therapy requires the isolation and purification of patient cells, the introduction of a therapeutic gene and introduction of the genetically altered cells back into the patient. In contrast, in vivo gene therapy does not require isolation and purification of a patient's cells.

The therapeutic gene is typically "packaged" for administration to a patient. Gene delivery vehicles may be non-viral, such as liposomes, or replication-deficient viruses, such as adenovirus as described by Berkner, K.L., in *Curr. Top. Microbiol. Immunol.*, 158, 39-66 (1992) or adeno-associated virus (AAV) vectors as described by Muzyczka, N., in *Curr. Top. Microbiol. Immunol.*, 158, 97-129 (1992) and U.S. Patent No. 5,252,479. For example, a nucleic acid molecule encoding a polypeptide of the invention

may be engineered for expression in a replication-defective retroviral vector. This expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding the polypeptide, such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells in vivo and expression of the polypeptide in vivo (see Chapter 20, Gene Therapy and other Molecular Genetic-based Therapeutic Approaches, (and references cited therein) in Human Molecular Genetics (1996), T Strachan and A P Read, BIOS Scientific Publishers Ltd).

- 10 Another approach is the administration of "naked DNA" in which the therapeutic gene is directly injected into the bloodstream or muscle tissue.

In situations in which the polypeptides or nucleic acid molecules of the invention are disease causing agents, the invention provides that they can be used in vaccines to raise antibodies against the disease causing agent.

- 15 Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection). Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with pharmaceutically-acceptable carriers as described above, which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, H. pylori, and other pathogens.

- 20 Since polypeptides may be broken down in the stomach, vaccines comprising polypeptides are preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents.

The vaccine formulations of the invention may be presented in unit-dose or multi-dose containers. For example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The dosage will depend on the specific activity of the vaccine and can be readily
5 determined by routine experimentation.

This invention also relates to the use of nucleic acid molecules according to the present invention as diagnostic reagents. Detection of a mutated form of the gene characterised by the nucleic acid molecules of the invention which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or
10 susceptibility to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques.

Nucleic acid molecules for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used
15 directly for detection or may be amplified enzymatically by using PCR, ligase chain reaction (LCR), strand displacement amplification (SDA), or other amplification techniques (see Saiki et al., *Nature*, 324, 163-166 (1986); Bej, et al., *Crit. Rev. Biochem. Molec. Biol.*, 26, 301-334 (1991); Birkenmeyer et al., *J. Virol. Meth.*, 35, 117-126 (1991); Van Brunt, J., *Bio/Technology*, 8, 291-294 (1990)) prior to analysis.

20 In one embodiment, this aspect of the invention provides a method of diagnosing a disease in a patient, comprising assessing the level of expression of a natural gene encoding a polypeptide according to the invention and comparing said level of expression to a control level, wherein a level that is different to said control level is indicative of disease. The method may comprise the steps of:

- 25 a)contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule of the invention and the probe;
- b)contacting a control sample with said probe under the same conditions used in step a);
- c)and detecting the presence of hybrid complexes in said samples;

Wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

A further aspect of the invention comprises a diagnostic method comprising the steps of:

- a) obtaining a tissue sample from a patient being tested for disease;
- 5 b) isolating a nucleic acid molecule according to the invention from said tissue sample; and
- c) diagnosing the patient for disease by detecting the presence of a mutation in the nucleic acid molecule which is associated with disease.

10 To aid the detection of nucleic acid molecules in the above-described methods, an amplification step, for example using PCR, may be included.

Deletions and insertions can be detected by a change in the size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labelled RNA of the invention or alternatively, labelled antisense DNA sequences of the invention. Perfectly matched sequences can be distinguished from
15 mismatched duplexes by RNase digestion or by assessing differences in melting temperatures. The presence or absence of the mutation in the patient may be detected by contacting DNA with a nucleic acid probe that hybridises to the DNA under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion
20 corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation in the corresponding portion of the DNA strand.

Such diagnostics are particularly useful for prenatal and even neonatal testing.

25 Point mutations and other sequence differences between the reference gene and "mutant" genes can be identified by other well-known techniques, such as direct DNA sequencing or single-strand conformational polymorphism, (see Orita et al., Genomics, 5, 874-879 (1989)). For example, a sequencing primer may be used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabelled

nucleotides or by automatic sequencing procedures with fluorescent-tags. Cloned DNA segments may also be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. Further, point mutations and other sequence variations, such as polymorphisms, can be detected as described above, for example, through the use of allele-specific oligonucleotides for PCR amplification of sequences that differ by single nucleotides.

DNA sequence differences may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (for example, Myers et al., *Science* (1985) 230: 1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (see Cotton et al., *Proc. Natl. Acad. Sci. USA* (1985) 85: 4397-4401).

In addition to conventional gel electrophoresis and DNA sequencing, mutations such as microdeletions, aneuploidies, translocations, inversions, can also be detected by in situ analysis (see, for example, Keller et al., *DNA Probes*, 2nd Ed., Stockton Press, New York, N.Y., USA (1993)), that is, DNA or RNA sequences in cells can be analysed for mutations without need for their isolation and/or immobilisation onto a membrane. Fluorescence in situ hybridization (FISH) is presently the most commonly applied method and numerous reviews of FISH have appeared (see, for example, Trachuck et al., *Science*, 250, 559-562 (1990), and Trask et al., *Trends, Genet.*, 7, 149-154 (1991)).

In another embodiment of the invention, an array of oligonucleotide probes comprising a nucleic acid molecule according to the invention can be constructed to conduct efficient screening of genetic variants, mutations and polymorphisms. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability (see for example: M.Chee et al., *Science* (1996), Vol 274, pp 610-613).

In one embodiment, the array is prepared and used according to the methods described in PCT application WO95/11995 (Chee et al); Lockhart, D. J. et al. (1996) *Nat. Biotech.* 14: 1675-1680; and Schena, M. et al. (1996) *Proc. Natl. Acad. Sci.* 93: 10614-10619). Oligonucleotide pairs may range from two to over one million. The oligomers are

synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support. In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al). In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536 or 6144 oligonucleotides, or any other number between two and over one million which lends itself to the efficient use of commercially-available instrumentation.

In addition to the methods discussed above, diseases may be diagnosed by methods comprising determining, from a sample derived from a subject, an abnormally decreased or increased level of polypeptide or mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification, for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

Assay techniques that can be used to determine levels of a polypeptide of the present invention in a sample derived from a host are well-known to those of skill in the art and are discussed in some detail above (including radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays). This aspect of the invention provides a diagnostic method which comprises the steps of: (a) contacting a ligand as described above with a biological sample under conditions suitable for the formation of a ligand-polypeptide complex; and (b) detecting said complex.

Protocols such as ELISA, RIA, and FACS for measuring polypeptide levels may additionally provide a basis for diagnosing altered or abnormal levels of polypeptide expression. Normal or standard values for polypeptide expression are established by

combining body fluids or cell extracts taken from normal mammalian subjects, preferably humans, with antibody to the polypeptide under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, such as by photometric means.

- 5 Antibodies which specifically bind to a polypeptide of the invention may be used for the diagnosis of conditions or diseases characterised by expression of the polypeptide, or in assays to monitor patients being treated with the polypeptides, nucleic acid molecules, ligands and other compounds of the invention. Antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics.
- 10 Diagnostic assays for the polypeptide include methods that utilise the antibody and a label to detect the polypeptide in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labelled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules known in the art may be used, several of which are described above.
- 15 Quantities of polypeptide expressed in subject, control and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease. Diagnostic assays may be used to distinguish between absence, presence, and excess expression of polypeptide and to monitor regulation of polypeptide levels during therapeutic intervention. Such assays
- 20 may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials or in monitoring the treatment of an individual patient.

A diagnostic kit of the present invention may comprise:

- (a) a nucleic acid molecule of the present invention;
- (b) a polypeptide of the present invention; or
- 25 (c) a ligand of the present invention.

In one aspect of the invention, a diagnostic kit may comprise a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to the invention; a second container containing primers useful for amplifying the nucleic acid molecule; and instructions for using the probe and primers for

facilitating the diagnosis of disease. The kit may further comprise a third container holding an agent for digesting unhybridised RNA.

In an alternative aspect of the invention, a diagnostic kit may comprise an array of nucleic acid molecules, at least one of which may be a nucleic acid molecule according to the invention.

To detect polypeptide according to the invention, a diagnostic kit may comprise one or more antibodies that bind to a polypeptide according to the invention; and a reagent useful for the detection of a binding reaction between the antibody and the polypeptide.

Such kits will be of use in diagnosing a disease or susceptibility to disease, particularly cardiovascular disease, cancer, asthma, COPD and inflammatory diseases.

Various aspects and embodiments of the present invention will now be described in more detail by way of example, with particular reference to the ADS1, ADS2, ADS3, ADS4 and ADS5 polypeptides.

It will be appreciated that modification of detail may be made without departing from the scope of the invention.

Brief description of the Figures

Figure 1: This is the front end of the Biopendium™ Target Mining Interface. A search of the database is initiated using the PDB code "1LFA:A".

Figure 2A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1LFA:A. The arrow indicates leukocyte integrin, a typical adhesion molecule.

Figure 2B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1LFA:A. The arrow indicates AAC74854.1 (ADS1).

Figure 2C: Full list of forward PSI-BLAST results for the search using 1LFA:A. AAC74854.1 (ADS1) is not identified.

Figure 3: The Redundant Sequence Display results page for AAC74854.1 (ADS1).

Figure 4: PFAM search results for AAC74854.1 (ADS1).

Figure 5: NCBI protein report for AAC74854.1 (ADS1).

Figure 6A: This is the front end of the Biopendium™ database. A search of the database is initiated using AAC74854.1 (ADS1), as the query sequence.

Figure 6B: A selection of the Inpharmatica Genome Threader results of search using AAC74854.1 (ADS1), as the query sequence. The arrow points to 1LFA:A.

Figure 6C: A selection of the reverse-maximised PSI-BLAST results obtained using AAC74854.1 (ADS1), as the query sequence.

Figure 7: AlEye sequence alignment of AAC74854.1 (ADS1) and 1LFA:A.

Figure 8A: LigEye for 1LFA: A that illustrates the sites of interaction of the bound metal ion required for adhesion activity with the metal binding ligands of the MIDAS motif of *Homo Sapiens* Leukocyte Function Antigen 1, 1LFA: A

Figure 8B: iRasMol view of 1LFA: A, *Homo Sapiens* Leukocyte Function Antigen 1. The coloured balls represent the amino acids in *Homo Sapiens* Leukocyte Function Antigen 1 that are involved in the MIDAS motif and that are conserved in AAC74854.1 (ADS1).

Figure 9: This is the front end of the Biopendium™ Target Mining Interface. A search of the database is initiated using the PDB code "1AOX:A".

Figure 10A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1AOX:A. The arrow indicates leukocyte integrin, a typical adhesion molecule.

Figure 10B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1AOX:A. The arrow indicates AAC76768.1 (ADS2).

Figure 10C: Full list of forward PSI-BLAST results for the search using 1AOX:A. AAC76768.1 (ADS2) is not identified.

Figure 11: The Redundant Sequence Display results page for AAC76768.1 (ADS2).

Figure 12: PFAM search results for AAC76768.1 (ADS2).

Figure 13: NCBI protein report for AAC76768.1 (ADS2).

Figure 14A: This is the front end of the Biopendium™ database. A search of the database is initiated using AAC76768.1 (ADS2), as the query sequence.

Figure 14B: A selection of the Inpharmatica Genome Threader results of search using AAC76768.1 (ADS2), as the query sequence. The arrow points to 1AOX:A.

- 5 Figure 14C: A selection of the reverse-maximised PSI-BLAST results obtained using AAC76768.1 (ADS2), as the query sequence.

Figure 15: AlEye sequence alignment of AAC76768.1 (ADS2) and 1AOX:A.

- Figure 16A: LigEye for 1AOX: A that illustrates the sites of interaction of the bound metal ion required for adhesion activity with the metal binding ligands of the MIDAS motif of *Homo Sapiens* Integrin Alpha 2 / Beta 1, 1AOX:A
- 10

Figure 16B: iRasMol view of 1AOX:A, *Homo Sapiens* Integrin Alpha 2 / Beta 1. The coloured balls represent the amino acids in *Homo Sapiens* Integrin Alpha 2 / Beta 1 that are involved in the MIDAS motif and that are conserved in AAC76768.1 (ADS2).

- Figure 17: This is the front end of the Biopendium™ Target Mining Interface. A search of the database is initiated using the PDB code "1JLM".
- 15

Figure 18A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1JLM. The arrow indicates leukocyte integrin, a typical adhesion molecule.

Figure 18B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1JLM. The arrow indicates AAD05645.1 (ADS3).

- 20 Figure 18C: Full list of forward PSI-BLAST results for the search using 1JLM. AAD05645.1 (ADS3) is not identified.

Figure 19: The Redundant Sequence Display results page for AAD05645.1 (ADS3).

Figure 20: PFAM search results for AAD05645.1 (ADS3).

Figure 21: NCBI protein report for AAD05645.1 (ADS3).

- 25 Figure 22A: This is the front end of the Biopendium™ database. A search of the database is initiated using AAD05645.1 (ADS3), as the query sequence.

Figure 22B: A selection of the Inpharmatica Genome Threader results of search using AAD05645.1 (ADS3), as the query sequence. The arrow points to 1JLM.

Figure 22C: A selection of the reverse-maximised PSI-BLAST results obtained using AAD05645.1 (ADS3), as the query sequence.

5 Figure 23: AIEye sequence alignment of AAD05645.1 (ADS3) and 1JLM.

Figure 24A: LigEye for 1JLM that illustrates the sites of interaction of the bound metal ion required for adhesion activity with the metal binding ligands of the MIDAS motif of *Homo Sapiens* Integrin CR3, 1JLM

10 Figure 24B: iRasMol view of 1JLM, *Homo Sapiens* Integrin CR3. The coloured balls represent the amino acids in *Homo Sapiens* Integrin CR3 that are involved in the MIDAS motif and that are conserved in AAD05645.1 (ADS3).

Figure 25: This is the front end of the Biopendium™ Target Mining Interface. A search of the database is initiated using the PDB code "1QC5:A".

15 Figure 26A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1QC5:A. The arrow indicates leukocyte integrin, a typical adhesion molecule.

Figure 26B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1QC5:A. The arrow indicates CAA57766.1 (ADS4).

20 Figure 26C: Full list of forward PSI-BLAST results for the search using 1QC5:A. CAA57766.1 (ADS4) is not identified.

Figure 27: The Redundant Sequence Display results page for CAA57766.1 (ADS4).

Figure 28: PFAM search results for CAA57766.1 (ADS4).

Figure 29: NCBI protein report for CAA57766.1 (ADS4).

25 Figure 30A: This is the front end of the Biopendium™ database. A search of the database is initiated using CAA57766.1 (ADS4), as the query sequence.

Figure 30B: A selection of the Inpharmatica Genome Threader results of search using CAA57766.1 (ADS4), as the query sequence. The arrow points to 1QC5:A.

Figure 30C: A selection of the reverse-maximised PSI-BLAST results obtained using CAA57766.1 (ADS4), as the query sequence.

Figure 31: ALEye sequence alignment of CAA57766.1 (ADS4) and 1QC5:A.

Figure 32A: LigEye for 1QC5:A that illustrates the sites of interaction of the bound metal ion required for adhesion activity with the metal binding ligands of the MIDAS motif of *Homo Sapiens* Integrin Alpha 1 / Beta 1, 1QC5: A

Figure 32B: iRasMol view of 1QC5:A, *Homo Sapiens* Integrin Alpha 1 / Beta 1. The coloured balls represent the amino acids in *Homo Sapiens* Integrin Alpha 1 / Beta 1 that are involved in the MIDAS motif and that are conserved in CAA57766.1 (ADS4).

Figure 33: This is the front end of the Biopendium™ Target Mining Interface. A search of the database is initiated using the PDB code “1JLM”.

Figure 34A: A selection is shown of the Inpharmatica Genome Threader results for the search using 1JLM. The arrow indicates leukocyte integrin, a typical adhesion molecule.

Figure 34B: A selection is shown of the Inpharmatica Genome Threader results for the search using 1JLM. The arrow indicates P10155 (ADS5).

Figure 34C: Full list of forward PSI-BLAST results for the search using 1JLM. P10155 (ADS5) is not identified.

Figure 35: The Redundant Sequence Display results page for P10155 (ADS5).

Figure 36: PFAM search results for P10155 (ADS5).

Figure 37: NCBI protein report for P10155 (ADS5).

Figure 38A: This is the front end of the Biopendium™ database. A search of the database is initiated using P10155 (ADS5), as the query sequence.

Figure 38B: A selection of the Inpharmatica Genome Threader results of search using P10155 (ADS5), as the query sequence. The arrow points to 1JLM.

Figure 38C: A selection of the reverse-maximised PSI-BLAST results obtained using P10155 (ADS5), as the query sequence.

Figure 39: ALEye sequence alignment of P10155 (ADS5) and 1JLM.

Figure 40A: LigEye for 1JLM that illustrates the sites of interaction of the bound metal ion required for adhesion activity with the metal binding ligands of the MIDAS motif of *Homo Sapiens* Integrin CR3, 1JLM

Figure 40B: iRasMol view of 1JLM, *Homo Sapiens* Integrin CR3. The coloured balls represent the amino acids in *Homo Sapiens* Integrin CR3 that are involved in the MIDAS motif and that are partly conserved in P10155 (ADS5).

Figure 41: AlEye sequence alignment of P10155, *Homo Sapiens* Ro60 (ADS5), and the *Mus musculus* (AAF19049.1), *Xenopus Laevis* (AAC38001.1) and *Caenorhabditis elegans* (CAA98241.1) Ro60 homologs.

Examples

Example 1:AAC74854.1 (ADS1)

In order to initiate a search for novel, distantly related adhesion molecules, an archetypal family member is chosen, the I-domain from *Homo Sapiens* Leukocyte Function Antigen 1. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the I-domain from *Homo Sapiens* Leukocyte Function Antigen 1 (PDB code 1LFA:A; see Figure 1).

A search of the Biopendium™ (using the Target Mining Interface) for relatives of 1LFA:A takes place and returns 2729 Genome Threader results. The 2729 Genome Threader results include examples of typical adhesion molecules, such as leukocyte integrin alpha chain (see arrow in Figure 2A).

Among the known adhesion molecules appears a protein of apparently unknown function, AAC74854.1 (ADS1; see arrow in figure 2B). The Inpharmatica Genome Threader has identified a sequence, AAC74854.1 (ADS1), as having a structure similar to *Homo Sapiens* Leukocyte Function Antigen 1, an adhesion molecule. The possession of a structure similar to an adhesion molecule suggests that AAC74854.1 (ADS1) functions as an adhesion molecule. The Genome Threader identifies this with 95% confidence.

The search of the Biopendium™ (using the Target Mining Interface) for relatives of 1LFA:A also returns 630 Forward PSI-Blast results. Forward PSI-Blast (see figure 2C) is unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify AAC74854.1 (ADS1) as an adhesion molecule.

- 5 In order to assess what is known in the public domain databases about AAC74854.1 (ADS1) the Redundant Sequence Display Page (Figure 3) is viewed. There are no associated PROSITE or PRINTS hits for AAC74854.1 (ADS1). PROSITE and PRINTS are databases that help to describe proteins of similar families. Returning no hits from both databases means that AAC74854.1 (ADS1) is unidentifiable as an adhesion
10 molecule using PROSITE or PRINTS. The redundant sequence display also shows any predicted features of AAC50543.1 (CCS5). These include potential coiled coil and low complexity regions in the sequence.

- In order to identify if any other public domain annotation vehicle is able to annotate AAC74854.1 (ADS1) as an adhesion molecule, the AAC74854.1 (ADS1) protein
15 sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models) (see Figure 4). The results identifies one PFAM-B match to AAC74854.1, however PFAM-B matches confer no functional annotation, only sequence similarity to other functionally unannotated proteins. Thus PFAM does not identify AAC74854.1 (ADS1) as an adhesion molecule.

- 20 The National Center for Biotechnology Information (NCBI) Genbank protein database is then viewed to examine if there is any further information that is known in the public domain relating to AAC74854.1 (ADS1). This is the U.S. public domain database for protein and gene sequence deposition (Figure 5). AAC74854.1 (ADS1) is a *Escherichia Coli* sequence, its Genbank protein ID is AAC74854.1 and it is 427 amino acids in
25 length. AAC74854.1 (ADS1) was cloned by a group of scientists at the University of Wisconsin, U.S.A. The entry identifies AAC74854.1 (ADS1) as a hypothetical protein. The public domain information for this gene does not annotate it as an adhesion molecule.

Therefore, it can be concluded that using all public domain annotation tools, AAC74854.1 (ADS1) may not be annotated as an adhesion molecule. Only the Inpharmatica Genome Threader is able to annotate this protein as an adhesion molecule.

5 The reverse search is now carried out. AAC74854.1 (ADS1) is now used as the query sequence in the Biopendium™ (see figure 6A). The Inpharmatica Genome Threader identifies AAC74854.1 (ADS1) as having a structure that is the same as *Homo Sapiens* Leukocyte Function Antigen 1 with 95% confidence (see arrow in figure 6B). *Homo Sapiens* Leukocyte Function Antigen 1 (1LFA) was the original query sequence. Positive iterations of PSI-Blast do not return this result (Figure 6C). It is only the Inpharmatica
10 Genome Threader that is able to identify this relationship.

The *Homo Sapiens* Leukocyte Function Antigen 1 sequence is chosen against which to view the sequence alignment of AAC74854.1 (ADS1). Viewing the AIEye alignment (Figure 7) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

15 The Leukocyte Function Antigen 1 I domain requires a bound metal ion in order to function. The metal ion forms a Metal Ion-Dependent Adhesion Site (MIDAS) which is characterised by a MIDAS motif consisting of the conserved metal liganding residues. The MIDAS motif in 1LFA:A consists of ASP10, SER12, SER14, THR79 and ASP112 all these residues are conserved in AAC74854.1 (ADS1) as ASP256, SER258, SER260,
20 THR315 and ASP346 respectively. The two serines and ASP112 are the metal ion ligands. This indicates that AAC74854.1 (ADS1) is an adhesion molecule similar to Leukocyte Function Antigen 1.

In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs “LigEye” (Figure 8A) and “iRasmol” (Figure 8B) are used. These
25 visualization tools identify the metal binding site of known protein structures by indicating the amino acids with which known metal ions or small molecule inhibitors interact at the active site. These interactions are through either a direct hydrogen bond or through hydrophobic interactions. In this manner one can see if the active site fold/structure is conserved between the identified homologue and the chosen protein of
30 known structure.

Since the structure of *Homo Sapiens* Leukocyte Function Antigen 1 is known (1LFA), this is chosen to illustrate the MIDAS motif (Figure 8B). ASP10, SER12, SER14, THR79 and ASP112 of 1LFA:A align with ASP256, SER258, SER260, THR315 and ASP346 of AAC74854.1. This indicates that indeed as predicted by the Inpharmatica Genome Threader, AAC74854.1 (ADS1) folds in a similar manner to *Homo Sapiens* Leukocyte Function Antigen 1 and as such is identified as an adhesion molecule.

Example 2: AAC76768.1 (ADS2)

In order to initiate a search for novel, distantly related adhesion molecules, an archetypal family member is chosen, the I-domain from *Homo Sapiens* Integrin Alpha 2 / Beta 1. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the I-domain from *Homo Sapiens* Integrin Alpha 2 / Beta 1 (PDB code 1AOX; see Figure 9).

A search of the Biopendium™ (using the Target Mining Interface) for relatives of 1AOX takes place and returns 2394 Genome Threader results. The 2394 Genome Threader results include examples of typical adhesion molecules, such as Integrin alpha 11 (see arrow in Figure 10A).

Among the known adhesion molecules appears a protein of apparently unknown function, AAC76768.1 (ADS2; see arrow in figure 10B). The Inpharmatica Genome Threader has identified a sequence, AAC76768.1 (ADS2), as having a structure similar to *Homo Sapiens* Integrin Alpha 2 / Beta 1, an adhesion molecule. The possession of a structure similar to an adhesion molecule suggests that AAC76768.1 (ADS2) functions as an adhesion molecule. The Genome Threader identifies this with 100% confidence.

The search of the Biopendium™ (using the Target Mining Interface) for homologues of 1FBL also returns 24 Reverse PSI-Blast results. The Inpharmatica Reverse PSI-Blast identifies AAC76768.1 (ADS2) as being related in sequence to *Homo Sapiens* Integrin Alpha 2 / Beta 1, detected in the -4 iteration (see Figure 10B, circled). The possession of a sequence related to an adhesion molecule suggests that AAC76768.1 (ADS2) functions as an adhesion molecule. This second proprietary method result consolidates the *Homo*

Sapiens Integrin Alpha 2 / Beta 1 structural relationship demonstrated with Genome Threader.

The search of the Biopendium™ (using the Target Mining Interface) for relatives of 1AOX also returns 608 Forward PSI-Blast results. Forward PSI-Blast (see figure 10C) is
5 unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify AAC76768.1 (ADS2) as an adhesion molecule.

In order to assess what is known in the public domain databases about AAC76768.1 (ADS2) the Redundant Sequence Display Page (Figure 11) is viewed. There are no associated PROSITE or PRINTS hits for AAC76768.1 (ADS2). PROSITE and PRINTS
10 are databases that help to describe proteins of similar families. Returning no hits from both databases means that AAC76768.1 (ADS2) is unidentifiable as an adhesion molecule using PROSITE or PRINTS. The redundant sequence display also shows any predicted features of AAC50543.1 (CCS5). These include a potential coiled coil in the sequence.

15 In order to identify if any other public domain annotation vehicle is able to annotate AAC76768.1 (ADS2) as an adhesion molecule, the AAC76768.1 (ADS2) protein sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models) (see Figure 12). The results identifies one PFAM-B match to AAC76768.1, however PFAM-B matches confer no functional annotation, only sequence
20 similarity to other functionally unannotated proteins. Thus PFAM does not identify AAC76768.1 (ADS2) as an adhesion molecule.

The National Center for Biotechnology Information (NCBI) Genbank protein database is then viewed to examine if there is any further information that is known in the public domain relating to AAC76768.1 (ADS2). This is the U.S. public domain database for
25 protein and gene sequence deposition (Figure 13). AAC76768.1 (ADS2) is a *Homo Sapiens* sequence, its Genbank protein ID is AAC76768.1 and it is 427 amino acids in length. AAC76768.1 (ADS2) was cloned by a group of scientists at the University of Wisconsin, USA. The entry identifies AAC76768.1 (ADS2) as a hypothetical protein. The public domain information for this gene does not annotate it as an adhesion
30 molecule.

Therefore, it can be concluded that using all public domain annotation tools, AAC76768.1 (ADS2) may not be annotated as an adhesion molecule. Only the Inpharmatica Genome Threader is able to annotate this protein as an adhesion molecule.

The reverse search is now carried out. AAC76768.1 (ADS2) is now used as the query sequence in the Biopendium™ (see figure 14A). The Inpharmatica Genome Threader identifies AAC76768.1 (ADS2) as having a structure that is the same as *Homo Sapiens* Integrin Alpha 2 / Beta 1 with 100% confidence (see arrow in figure 14B). *Homo Sapiens* Integrin Alpha 2 / Beta 1 (1AOX) was the original query sequence. The first 3 iterations of positive PSI-Blast do not return this result (Figure 14C), adhesion molecules are only detected at and above iteration 4. It is only the Inpharmatica Genome Threader that is able to identify this relationship.

The *Homo Sapiens* Integrin Alpha 2 / Beta 1 sequence is chosen against which to view the sequence alignment of AAC76768.1 (ADS2). Viewing the AIEye alignment (Figure 15) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

The Integrin Alpha 2 / Beta 1 I domain requires a bound metal ion in order to function. The metal ion forms a Metal Ion-Dependent Adhesion Site (MIDAS) which is characterised by a MIDAS motif consisting of the conserved metal liganding residues. The MIDAS motif in 1AOX consists of ASP13, SER15, SER17, THR83 and ASP116 all these residues are conserved in AAC76768.1 (ADS2) as ASP271, SER273, SER275, THR337 and ASP365 respectively. The two serines and ASP116 are the metal ion ligands. This indicates that AAC76768.1 (ADS2) is an adhesion molecule similar to Integrin Alpha 2 / Beta 1.

In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs “LigEye” (Figure 16A) and “iRasmol” (Figure 16B) are used. These visualization tools identify the metal binding site of known protein structures by indicating the amino acids with which known metal ions or small molecule inhibitors interact at the active site. These interactions are through either a direct hydrogen bond or through hydrophobic interactions. In this manner one can see if the active site

fold/structure is conserved between the identified homologue and the chosen protein of known structure.

Since the structure of *Homo Sapiens* Integrin Alpha 2 / Beta 1 is known (1AOX), this is chosen to illustrate the MIDAS motif (Figure 16B). ASP13, SER15, SER17, THR83 and ASP116 of 1AOX align with ASP271, SER273, SER275, THR337 and ASP365 of AAC76768.1. This indicates that indeed as predicted by the Inpharmatica Genome Threader, AAC76768.1 (ADS2) folds in a similar manner to *Homo Sapiens* Integrin Alpha 2 / Beta 1 and as such is identified as an adhesion molecule.

Example 3:AAD05645.1 (ADS3)

In order to initiate a search for novel, distantly related adhesion molecules, an archetypal family member is chosen, the I-domain from *Homo Sapiens* Integrin CR3. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the I-domain from *Homo Sapiens* Integrin CR3 (PDB code 1JLM; see Figure 17).

A search of the Biopendium™ (using the Target Mining Interface) for relatives of 1JLM takes place and returns 2925 Genome Threader results. The 2925 Genome Threader results include examples of typical adhesion molecules, such as leukocyte integrin (see arrow in Figure 18A).

Among the known adhesion molecules appears a protein of apparently unknown function, AAD05645.1 (ADS7; see arrow in figure 18B). The Inpharmatica Genome Threader has identified a sequence, AAD05645.1 (ADS3), as having a structure similar to *Homo Sapiens* Integrin CR3, an adhesion molecule. The possession of a structure similar to an adhesion molecule suggests that AAD05645.1 (ADS3) functions as an adhesion molecule. The Genome Threader identifies this with 89% confidence.

The search of the Biopendium™ (using the Target Mining Interface) for relatives of 1JLM also returns 626 Forward PSI-Blast results. Forward PSI-Blast (see figure 18C) is unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify AAD05645.1 (ADS3) as an adhesion molecule.

In order to assess what is known in the public domain databases about AAD05645.1 (ADS3) the Redundant Sequence Display Page (Figure 19) is viewed. There are no associated PROSITE or PRINTS hits for AAD05645.1 (ADS3). PROSITE and PRINTS are databases that help to describe proteins of similar families. Returning no hits from both databases means that AAD05645.1 (ADS3) is unidentifiable as an adhesion molecule using PROSITE or PRINTS. The redundant sequence display also shows any predicted features of AAC50543.1 (CCS5). These include a potential coiled coil region in the middle of the sequence.

In order to identify if any other public domain annotation vehicle is able to annotate AAD05645.1 (ADS3) as an adhesion molecule, the AAD05645.1 (ADS3) protein sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models) (see Figure 20A). The results identifies two PFAM-B matches to AAD05645.1, however PFAM-B matches confer no functional annotation, only sequence similarity to other functionally unannotated proteins. The result also identifies one PFAM-A match, however this is to the FtsK/SpoIIIE family (see Figure 20B) which does not confer any particular functional annotation. Thus PFAM does not identify AAD05645.1 (ADS3) as an adhesion molecule.

The National Center for Biotechnology Information (NCBI) Genbank protein database is then viewed to examine if there is any further information that is known in the public domain relating to AAD05645.1 (ADS3). This is the U.S. public domain database for protein and gene sequence deposition (Figure 21). AAD05645.1 (ADS3) is a *Homo Sapiens* sequence, its Genbank protein ID is AAD05645.1 and it is 806 amino acids in length. AAD05645.1 (ADS3) was cloned by a group of scientists at the Astra Research Center Boston in Cambridge, USA. The entry identifies AAD05645.1 (ADS3) as a putative protein. The public domain information for this gene does not annotate it as an adhesion molecule.

Therefore, it can be concluded that using all public domain annotation tools, AAD05645.1 (ADS3) may not be annotated as an adhesion molecule. Only the Inpharmatica Genome Threader is able to annotate this protein as an adhesion molecule.

The reverse search is now carried out. AAD05645.1 (ADS7) is now used as the query sequence in the Biopendium™ (see figure 22A). The Inpharmatica Genome Threader identifies AAD05645.1 (ADS3) as having a structure that is the same as *Homo Sapiens* Integrin CR3 with 89% confidence (see arrow in figure 22B). *Homo Sapiens* Integrin CR3 (1JLM) was the original query sequence. Positive iterations of PSI-Blast do not return this result (Figure 22C). It is only the Inpharmatica Genome Threader that is able to identify this relationship.

The *Homo Sapiens* Integrin CR3 sequence is chosen against which to view the sequence alignment of AAD05645.1 (ADS3). Viewing the AIEye alignment (Figure 23) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

The Integrin CR3 I domain requires a bound metal ion in order to function. The metal ion forms a Metal Ion-Dependent Adhesion Site (MIDAS) which is characterised by a MIDAS motif consisting of the conserved metal liganding residues. The MIDAS motif in 1JLM consists of ASP9, SER11, SER13, THR78 and ASP111 all these residues are conserved in AAD05645.1 (ADS3) as ASP350, SER352, SER354, THR415 and ASP464 respectively. The two serines and ASP111 are the metal ion ligands. This indicates that AAD05645.1 (ADS3) is an adhesion molecule similar to Integrin CR3.

In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs “LigEye” (Figure 24A) and “iRasmol” (Figure 24B) are used. These visualization tools identify the metal binding site of known protein structures by indicating the amino acids with which known metal ions or small molecule inhibitors interact at the active site. These interactions are through either a direct hydrogen bond or through hydrophobic interactions. In this manner one can see if the active site fold/structure is conserved between the identified homologue and the chosen protein of known structure.

Since the structure of *Homo Sapiens* Integrin CR3 is known (1JLM), this is chosen to illustrate the MIDAS motif (Figure 24B). ASP9, SER11, SER13, THR78 and ASP111 of 1JLM align with ASP350, SER352, SER354, THR415 and ASP464 of AAD05645.1.

This indicates that indeed as predicted by the Inpharmatica Genome Threader,

AAD05645.1 (ADS3) folds in a similar manner to *Homo Sapiens* Integrin CR3 and as such is identified as an adhesion molecule.

Example 4:VP4 (ADS4)

In order to initiate a search for novel, distantly related adhesion molecules, an archetypal family member is chosen, the I-domain from *Homo Sapiens* Integrin Alpha 1 / Beta 1. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the I-domain from *Homo Sapiens* Integrin Alpha 1 / Beta 1 (PDB code 1QC5; see Figure 25).

A search of the Biopendium™ (using the Target Mining Interface) for relatives of 1QC5:A takes place and returns 3331 Genome Threader results. The 3331 Genome Threader results include examples of typical adhesion molecules, such as leukocyte integrin (see arrow in Figure 26A).

Among the known adhesion molecules appears a protein of apparently unknown function, VP4 (ADS4; see arrow in figure 26B). The Inpharmatica Genome Threader has identified a sequence, VP4 (ADS4), as having a structure similar to *Homo Sapiens* Integrin Alpha 1 / Beta 1, an adhesion molecule. The possession of a structure similar to an adhesion molecule suggests that VP4 (ADS4) functions as an adhesion molecule. The Genome Threader identifies this with 48% confidence.

The search of the Biopendium™ (using the Target Mining Interface) for relatives of 1QC5:A also returns 622 Forward PSI-Blast results. Forward PSI-Blast (see figure 26C) is unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify VP4 (ADS4) as an adhesion molecule.

In order to assess what is known in the public domain databases about VP4 (ADS4) the Redundant Sequence Display Page (Figure 27) is viewed. There are no associated PROSITE or PRINTS hits for VP4 (ADS4). PROSITE and PRINTS are databases that help to describe proteins of similar families. Returning no hits from both databases means that VP4 (ADS4) is unidentifiable as an adhesion molecule using PROSITE or PRINTS.

The redundant sequence display also shows any predicted features of CAA57766.1 (ADS4). These include a potential coiled coil region at the start of the sequence.

In order to identify if any other public domain annotation vehicle is able to annotate VP4 (ADS4) as an adhesion molecule, the VP4 (ADS4) protein sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models) (see Figure 28). The results identifies one PFAM-A match to VP4, however the PFAM-A match is to the VP4 domain, which only annotates the protein as a VP4 protein and confers no additional functional annotation. Thus PFAM does not identify VP4 (ADS4) as an adhesion molecule.

The National Center for Biotechnology Information (NCBI) Genbank protein database is then viewed to examine if there is any further information that is known in the public domain relating to VP4 (ADS4). This is the U.S. public domain database for protein and gene sequence deposition (Figure 29). VP4 (ADS4) is a *Human Rotavirus* sequence, its Genbank protein ID is CAA57766.1 and it is 775 amino acids in length. VP4 (ADS4) was cloned by a group of scientists at the Indian Institute of Science, Bangalore, India. The entry identifies VP4 (ADS4) as a coat protein of *Human Rotavirus*. The public domain information for this gene does not annotate it as an adhesion molecule.

Therefore, it can be concluded that using all public domain annotation tools, VP4 (ADS4) may not be annotated as an adhesion molecule. Only the Inpharmatica Genome Threader is able to annotate this protein as an adhesion molecule.

The reverse search is now carried out. CAA57766.1 (ADS4) is now used as the query sequence in the Biopendium™ (see figure 30A). The Inpharmatica Genome Threader identifies VP4 (ADS4) as having a structure that is the same as *Homo Sapiens* Integrin Alpha 1 / Beta 1 with 48% confidence (see arrow in figure 30B). *Homo Sapiens* Integrin Alpha 1 / Beta 1 (1QC5) was the original query sequence. The Von Willebrand domains also hit by VP4 are also adhesion molecules with a very similar structure to the integrin I domain. Positive iterations of PSI-Blast do not return this result (Figure 30C). It is only the Inpharmatica Genome Threader that is able to identify this relationship.

The *Homo Sapiens* Integrin Alpha 1 / Beta 1 sequence is chosen against which to view the sequence alignment of VP4 (ADS4). Viewing the AIEye alignment (Figure 31) of the

query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

The Integrin Alpha 1 / Beta 1 I domain requires a bound metal ion in order to function. The metal ion forms a Metal Ion-Dependent Adhesion Site (MIDAS) which is characterised by a MIDAS motif consisting of the conserved metal liganding residues. The MIDAS motif in 1QC5:A consists of ASP11, SER13, SER15, THR81 and ASP114 all these residues are conserved in VP4 (ADS4) as ASP592, SER594, SER596, THR654 and GLU686 respectively. The change of ASP114 to GLU686 is a conservative change as both residues have acidic groups capable of binding metal. The two serines and ASP114 are the metal ion ligands. This indicates that VP4 (ADS4) is an adhesion molecule similar to Integrin Alpha 1 / Beta 1.

In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs “LigEye” (Figure 32A) and “iRasmol” (Figure 32B) are used. These visualization tools identify the metal binding site of known protein structures by indicating the amino acids with which known metal ions or small molecule inhibitors interact at the active site. These interactions are through either a direct hydrogen bond or through hydrophobic interactions. In this manner one can see if the active site fold/structure is conserved between the identified homologue and the chosen protein of known structure.

Since the structure of *Homo Sapiens* Integrin Alpha 1 / Beta 1 is known (1QC5), this is chosen to illustrate the MIDAS motif (Figure 63B). of ASP11, SER13, SER15, THR81 and ASP114 of 1QC5:A align with ASP592, SER594, SER596, THR654 and GLU686 of VP4. This indicates that indeed as predicted by the Inpharmatica Genome Threader, VP4 (ADS4) folds in a similar manner to *Homo Sapiens* Integrin Alpha 1 / Beta 1 and as such is identified as an adhesion molecule.

Example 5:Ro60 (ADS5)

In order to initiate a search for novel, distantly related adhesion molecules, an archetypal family member is chosen, the I-domain from *Homo Sapiens* Integrin CR3. More specifically, the search is initiated using a structure from the Protein Data Bank (PDB) which is operated by the Research Collaboratory for Structural Bioinformatics.

The structure chosen is the I-domain from *Homo Sapiens* Integrin CR3 (PDB code 1JLM; see Figure 33).

A search of the Biopendium™ (using the Target Mining Interface) for relatives of 1JLM takes place and returns 2925 Genome Threader results. The 2925 Genome Threader results include examples of typical adhesion molecules, such as leukocyte integrin (see arrow in Figure 34A).

Among the known adhesion molecules appears a protein of apparently unknown function, Ro60 (ADS5; see arrow in figure 34B). The Inpharmatica Genome Threader has identified a sequence, Ro60 (ADS5), as having a structure similar to *Homo Sapiens* Integrin CR3, an adhesion molecule. The possession of a structure similar to an adhesion molecule suggests that Ro60 (ADS5) functions as an adhesion molecule. The Genome Threader identifies this with 70% confidence.

The search of the Biopendium™ (using the Target Mining Interface) for relatives of 1JLM also returns 626 Forward PSI-Blast results. Forward PSI-Blast (see figure 34C) is unable to identify this relationship; only the Inpharmatica Genome Threader is able to identify Ro60 (ADS5) as an adhesion molecule.

In order to assess what is known in the public domain databases about Ro60 (ADS5) the Redundant Sequence Display Page (Figure 35) is viewed. There are no associated PROSITE or PRINTS hits for Ro60 (ADS5). PROSITE and PRINTS are databases that help to describe proteins of similar families. Returning no hits from both databases means that Ro60 (ADS5) is unidentifiable as an adhesion molecule using PROSITE or PRINTS. The redundant sequence display also shows any predicted features of Ro60 (ADS5). These include a potential coiled coil region at the start of the sequence and a transmembrane region. Although transmembrane regions are not predictive of adhesion molecules, they are a common characteristic of adhesion molecules. Thus the possession of a transmembrane region consolidates the Inpharmatica Genome Threader annotation of Ro60 (ADS5) as an adhesion molecule.

In order to identify if any other public domain annotation vehicle is able to annotate Ro60 (ADS5) as an adhesion molecule, the Ro60 (ADS5) protein sequence is searched against the PFAM database (Protein Family Database of Alignment and hidden Markov models)

(see Figure 36). The results identifies two PFAM-B matches to Ro60, however PFAM-B matches confer no functional annotation, only sequence similarity to other functionally unannotated proteins. Thus PFAM does not identify Ro60 (ADS5) as an adhesion molecule.

- 5 The National Center for Biotechnology Information (NCBI) Genbank protein database is then viewed to examine if there is any further information that is known in the public domain relating to Ro60 (ADS5). This is the U.S. public domain database for protein and gene sequence deposition (Figure 37). Ro60 (ADS5) is a *Homo Sapiens* sequence, its SWISS-PROT protein ID is P10155 and it is 538 amino acids in length. Ro60 (ADS5)
10 was cloned by a group of scientists at the W.M. Keck Autoimmune Disease Center, California. The entry identifies Ro60 (ADS5) as a RNA binding protein associated with the autoimmune disease: Sjogren's Syndrome. The public domain information for this gene does not annotate it as an adhesion molecule.

Therefore, it can be concluded that using all public domain annotation tools, Ro60
15 (ADS5) may not be annotated as an adhesion molecule. Only the Inpharmatica Genome Threader is able to annotate this protein as an adhesion molecule.

The reverse search is now carried out. P10155 (ADS5) is now used as the query sequence in the Biopendium™ (see figure 38A). The Inpharmatica Genome Threader identifies Ro60 (ADS5) as having a structure that is the same as *Homo Sapiens* Integrin CR3 with
20 70% confidence (see arrow in figure 38B). *Homo Sapiens* Integrin CR3 (1JLM) was the original query sequence. Positive iterations of PSI-Blast do not return this result (Figure 38C). It is only the Inpharmatica Genome Threader that is able to identify this relationship.

The *Homo Sapiens* Integrin CR3 sequence is chosen against which to view the sequence
25 alignment of Ro60 (ADS5). Viewing the AlEye alignment (Figure 39) of the query protein against the protein identified as being of a similar structure helps to visualize the areas of homology.

The Integrin CR3 I domain requires a bound metal ion in order to function. The metal ion forms a Metal Ion-Dependent Adhesion Site (MIDAS) which is characterised by a
30 MIDAS motif consisting of the conserved metal liganding residues. The MIDAS motif in

1JLM consists of ASP9, SER11, SER13, THR78 and ASP111 all these residues except for THR78 are conserved in Ro60 (ADS5) as ASP376, SER378, SER380, ILE440 and ASP469 respectively. The two serines and ASP111 are the metal ion ligands. This indicates that Ro60 (ADS5) is an adhesion molecule similar to Integrin CR3.

- 5 In order to ensure that the protein identified is in fact a relative of the query sequence, the visualization programs “LigEye” (Figure 40A) and “iRasmol” (Figure 40B) are used. These visualization tools identify the metal binding site of known protein structures by indicating the amino acids with which known metal ions or small molecule inhibitors interact at the active site. These interactions are through either a direct hydrogen bond or
10 through hydrophobic interactions. In this manner one can see if the active site fold/structure is conserved between the identified homologue and the chosen protein of known structure.

Since the structure of *Homo Sapiens* Integrin CR3 is known (1JLM), this is chosen to illustrate the MIDAS motif (Figure 63B). ASP9, SER11, SER13, THR78 and ASP111 of
15 1JLM align with ASP376, SER378, SER380, ILE440 and ASP469 of Ro60. This indicates that indeed as predicted by the Inpharmatica Genome Threader, Ro60 (ADS5) folds in a similar manner to *Homo Sapiens* Integrin CR3 and as such is identified as an adhesion molecule.

Reverse-maximised PSI-BLAST of Ro60 (ADS5) identifies *Mus musculus*, *Xenopus*
20 *Laevis* and *Caenorhabditis elegans* homologs of Ro60 (ADS5) called AAF19049.1, AAC38001.1, and CAA98241.1 respectively. AAF19049.1 has 90.0% sequence identity to P10155 (*Homo sapiens* Ro60;ADS5), see Figure 38C. AAC38001.1 has 76.0% sequence identity to P10155 (*Homo sapiens* Ro60;ADS5), see Figure 38C. CAA98241.1 has 36.0% sequence identity to P10155 (*Homo sapiens* Ro60;ADS5), see Figure 38C.

25 P10155 (*Homo sapiens* Ro60;ADS5), called AAF19049.1, AAC38001.1, and CAA98241.1 are aligned and viewed in AlEye (Figure 41). AlEye reveals that the predicted metal binding residues SER378, SER380 and ASP469 of P10155 (*Homo sapiens* Ro60;ADS5), are conserved in the *Mus musculus* (AAF19049.1), *Xenopus Laevis* (AAC38001.1) and *Caenorhabditis elegans* (CAA98241.1) DICE-1 homologs. Another
30 predicted MIDAS residue of (*Homo sapiens* Ro60;ADS5), ASP376, is also conserved.

Residues, which are essential for the function of a protein, will be conserved in homologs of that protein. Thus the conservation of SER378, SER380 and ASP469 (residues which would be essential for the function of the I-domain) and ASP376 in the *Mus musculus* (AAF19049.1), *Xenopus Laevis* (AAC38001.1) and *Caenorhabditis elegans* (CAA98241.1) DICE-1 homologs strongly supports the annotation of P10155 (*Homo sapiens* Ro60;ADS5) as an adhesion molecule.

Sequence Listing

SEQ ID NO: 1 Nucleotide coding sequence for AAC74854.1 (ADS1) protein

```

1  atgacctggt ttattgaccg gcgtctgaac ggcaaaaaca aaagcatggt gaatcgccag
61  cgtttttttac gccgttataa agcgcaaatt aaacagtcga tctccgaggc cattaataag
5  121  cgttcgggtga ctgacgtcga cagcggcgaa tccgtatcca ttcccacgga agatattagc
181  gaaccgatgt ttcacagggg gcgtggcggt ctgcgccacc gcgtgcatcc gggcaatgac
241  catttcgtcc agaacgaccg aattgaacgt ccccaggggtg gcggcggagg ttccggcagt
301  ggtcagggcc aggccagcca ggatggtgaa ggtcaggatg aatttgtctt tcagatttcg
361  aaagatgagt atcttgatct gctctttgaa gatttggcct taccgaatct gaaacaaaac
10 421  caacaacgcc agctgaccga atataaaacg catcggggcg gttataccgc taacggcggt
481  ccggccaata tcagcgttgt gcgttcattg cagaactcac tggcgcgacg cacagccatg
541  acggcaggca agcggcgagg acttcattgca ctggaagaga atttggccat catcagcaac
601  agtgaacctg cgcaactgct ggaagaggaa cgtctgcgca aagaaattgc agaattacgt
661  gccaaaattg aacgcgtccc ttttattgac accttcgatt tacgttacia gaactacgag
15 721  aagcggcccc atccctccag ccaggcagtg atgttttgcc tgatggacgt ttccgggtca
781  atggatcaat ccactaaaga tatggctaag cgtttttata ttctgctgta tctgttcctc
841  agcagaacgt ataagaacgt ggaagtcgta tacatccgcc atcataccca ggcgaaagaa
901  gtcgatgaac atgagttttt ctactcgcag gaaacaggcg gcaccattgt ttccagcgcc
961  ctgaaactga tggatgaggt agtgaaagag cgttataacc cggcacagtg gaatatttac
20 1021 gctgcacaag catcggacgg cgataactgg gccgatgact ctccgctttg ccatgaaatc
1081 ctggcgaaaa aattattacc tgttgttcgt tattacagct atatcgaaat taccgctcgt
1141 gcacatcaga cattgtggcg agaatatgag catctgcaat ctactttcga caactttgcg
1201 atgcagcaca tccgcgacca ggatgatatt tatccggtgt tccgtgaact gtttcataaa
1261 caaatgcaa cagctaaagg ctaa

```

25 SEQ ID NO: 2 Protein AAC74854.1; ADS1

```

1  mtwfidrrln gknksmvnrq rflrryqaqi kqsiseaink rsvtdvdsge svsiptedis
61  epmfhqgrgg lrhrvhpgnd hfvqndrier pggggggsgs gggqasqdge gqdefvfqis
121  kdeyldllfe dlalpnlkqn qqrqlteykt hragytangv panisvvrsl qnslarrtam
181  tagkrrelha leenlaiisn sepaqlleee rlrkeiaelr akiervpfid tfdlryknye

```

241 krpdpsqav mfcldvsgs mdqstkdma rfyillylfl srtyknvevv yirhhtqake
 301 vdeheffysq etggativssa lklddevvke rynpaqwniy aaqasdgdnw addsplchei
 361 laklllpvvr ysysieitrr ahqtlwreya hlqstfdnfa mqhirdqddi ypvfrelfhk
 421 qnatakq

5 SEQ ID NO: 3 Nucleotide coding sequence for AAC76768.1 (ADS2) protein

1 gtgcgcagtc ggctgaaaga tgcccgagtc ccgccggaac tcaccgaaga ggtgatgtgc
 61 tatcagcaaa gccagctcct ctccacgcca cagtttattg tgcagctacc acagatcctg
 121 gacttactgc atcgtctgaa ttctccatgg gcagaacaag cccgacagtt ggttgatgct
 181 aacagcacga tcaattcagc gttacacacg ctttttctcc agcgttggcg ttttaagtctg
 241 atcgtgcaag caacgacgtt aatcaacag ctattagaag aagaacgca acaactgttg
 301 agtgaagttc aggaacgcat gacgctgagc ggacaacttg aaccgattct cgcagataac
 361 aatactgcag ctggctgtct gtgggatatg agcgccggtc agcttaaacg tggcgactat
 421 cagttgattg tgaaatacgg tgaatttctt aacgaacagc cggaactgaa acgcttgga
 481 gagcagctgg ggcgttctcg ggaagccaaa tcaataccgc gcaacgatgc gcagatggaa
 541 accttccgca ccatggtgag cgaaccggcg acggttctcg agcaggttga tggctgcaa
 601 caaagcgatg atattttacg tctcctgccg ccagaactgg cgacactagg gataacggaa
 661 ctggagtatg agttttaccg tcggctggtg gaaaaacagt tgctcaccta tcgctgcac
 721 ggtgagtcgt ggcgtgaaaa agtgatcgaa cgtccgggtg tacataaaga ttacgatgaa
 781 cagccgcgag ggcgtttat tgtctgtgtg gatacttccg gctcaatggg cggctttaat
 841 gaacagtgtg cgaaagcgtt ctgcctggcc ttgatgca caaagacagg agaaaaccgg
 901 cgctgctata ttatgctatt ttccaccgag atcgtccggt atgagctttc aggccacaa
 961 ggcacgaac aagcaatccg ttttttaagc cagcagtttc gtggcggcac cgatcttgcc
 1021 agttgttttc gcgccattat ggaacgcttg caaagcaggg aatggtttga tgccgatgag
 1081 gtgggtgatt ctgattttat cgctcagcgg ttgcctgacg acgtgacgag taaagtga
 1141 gagctgcagc ggggtacatca gcacgcttt catgccgtgg cgatgtcggc acacggcaaa
 1201 ccggcatca tgcgcatttt cgatcatatc tggcgctttg ataccgggat gcgaagccgc
 1261 ctgctcagac gctggcgagg ataa

SEQ ID NO: 4 Protein AAC76768.1; ADS2

1 mrsrlkdarv ppelteevmc yqqsllstp qfivqlpqil dllhrlnspw aeqarqlvda

61 nstitsalht lflqrwrlsl ivqattlnqq lleereqll sevqermtls gglepiladn
 121 ntaagrlwdm sagqlkrgdy qlivkygefl neqpelkrla eqlgrsreak siprndaqme
 181 tfrtmvrepv tvpeqvdlgq qsddilrllp pelatlgite leyefyrrlv ekqllytyrlh
 241 geswrekvie rpvvhkdyde qprgpfivcv dtsgsmggfn eqcakafcla lmrialae
 5 301 rcyimlfste ivryelsgpq gieqairfls qqfrggttdla scfraimerl qsrewfdada
 361 vvisdfiaqr lpddvtskvk elqrvhqhrf havamsahgk pgimrifdhi wrfdtgmrsr
 421 llrrwrr

SEQ ID NO: 5 Nucleotide coding sequence for AAD05645.1 (ADS3) protein

1 atgattgaaa taagcgaatg gttgcaaaaa ctagacgatg ccttagataa agttgttgct
 10 61 aaaaaagagc cagagagttt tctcaagccg atcatttcac caatagagga ctacaaaaag
 121 agtgtcaggc aaattcaagc gcaattcaca gacgcgccga agttcaatga agaggggtgct
 181 taccctcaat ttttaagctg tggtttattg caagttaggg gcaaaaatgg tgctaacatg
 241 gaatttttat tgcctaaagt ttatcctttc cccctaaaa gcttgatat agagcatgaa
 301 aaagacgggc agtttttgag agaaatgctc atgcgcttac tctccagcgc gccttttagt
 15 361 caattggaag tgatcttaat tgatgcgttg agcttggggg gcattttcaa tctggccaga
 421 aggcttttag ataaaaacaa tgactttatt taccagcaaa ggattttgac cgaaagcaag
 481 gaaatagaag aagccctaaa gcatttgcat gaatatttaa aggttaattt gcaagaaaaa
 541 ttagccggtt ttagagattt tgtgcattat aatgaaaacg ccaaagactc cttgccttta
 601 aaagcgcttt ttttaagcgg ggtggatgct ttgagtaaag acgcgcttta ttatctagaa
 20 661 aagatcatgc gttttggctc taaaaatggg gttttgagct ttgtcaattt ggagagcgaa
 721 aaaaacaatc aatccgcaga agatttgaaa cgctatgcgg agttttttaa agacaggaca
 781 agttttgagt gcttaaaata ccttaatgta gaaatcatca gcgatcaagg tattaatcc
 841 caacacatgc aagacttcgc tgataaaatt aaagcgtatt acaagcaaaa aaaagaagtt
 901 aaaagggagt tgaaggactt acaaagagac aaagaatttt ggactaaaag ctctcagcat
 25 961 gaagtaagcg tgccggtggg gtgggatatt aaccataagg aagtgtgttt taaaatcggt
 1021 aacgaacaaa accacacgct catttgcgac cacagcggga gtgggaaatc caatttcttg
 1081 catgtgttga ttcaaaatct agctttctac tacgatcccg atgaagtcca actctttttg
 1141 ttagactata aagagggggg ggaatttaac gcgatgtag cagatccgc ttagagcat
 1201 gcgaggttgg tgagcgtggc gagttcaatc tcttatggca tcaacttctt gaaatggctt
 30 1261 tgtgatgaaa tgcaaaaaag agccgatcgg ttcaagcagt ttaatgtgaa agatttaagc

1321 gattaccgca aacatgaaaa aatgcccgaga ctgatcgtgg tgattgatga atttcagggtg
 1381 ctttttagcg ataataaatc cactaaagcg gtggaggggc atttaaacac cctgcttaaa
 1441 aagggccgta gctatgggggt gcatttggtt ttggccactc aaaccatgcg cggcactgac
 1501 attaatacaa gctttaaggg tcaaatcgcc aaccgcatcg ctttgcctat ggatgcagaa
 5 1561 gacagcagta gtgttttggg cgatgatgcg gcttgtgaga ttcaaaaacc agaaggcatt
 1621 ttcaacaaca acggagggaa tagaaaatac cacaccaaga tgagtgtccc taaagcccct
 1681 gatgatttca aatcttttct cacaaaaata cacgctgaat ttaaccaaag aaatctcgca
 1741 cccatagatc gtaaaatcta taatggcgag acacctttaa aaatgcccgga cacccttaag
 1801 gctaatagaaa tgcgtttgca tctgggcaaa aaagtggatt atgagcaaaa ggacctgata
 10 1861 gtggagtttg aaagtaacga atcgcatttg ttggtggtga tccaagattt aaacgctcgc
 1921 atcgctttta tgaaactctt attccaaaac gttaagagcg cgaacaaaga attgggtttt
 1981 tgcaataaag aaaaacgctt gataaggtct tttgatgcac aaaaagaata cggcatcacg
 2041 cctgtagaaa atattttaag cgtttttagac accgctatga atcctaacag cgcgcttggtg
 2101 atagacaatc tcaacgaagc gaaagaattg cagcaciaag taggggaggga aaagttaaaa
 15 2161 tcgttttttag aaaaagccat agacaacgag cagtattgcg tcatttttgc gcatgacttt
 2221 aggcaatta aaactaatta ccattttgac aagttaaaag aattgttaaa caaccacttc
 2281 aagcaatgcc tagccttttag gtgcaatggg gaaaacttga acgctatcaa aagcgatttg
 2341 cctccaccaa gcaaacttaa cgtgctattg atagagcttt ccaaagacag cgttactgaa
 2401 ttcaggcctt tcagcttata g

20 SEQ ID NO: 6 Protein AAD05645.1; ADS3

1 mieisewlqk lddaldkvva kkespflkp iispiedyqk svrqiaqft dapkfneega
 61 ypqlflscgll qvrgknganm eflpkvypf ppkslyiehe kdgqflreml mrlssaplv
 121 qlevilidal slggifnlar rlldknndfi yqgriltesk eieealkhlh eylknvlqek
 181 lagfrdfvhy nenakdslpl kalflsgvda lskdalyyle kimrfgskng vlsfvnlese
 25 241 knnqsaedlk ryaeffkdrf sfeclkylnv eiisdqgiks qhmqdfadki kayyqkqkev
 301 krelkdlqrd kefwtkssqh evsvpvgwdi nhkevcfkig neqnthtlicd hsgsgksnfl
 361 hvliqnlafy ydpdevqlfl ldykegvefn ayvadpaleh arlvsvassi sygitflklw
 421 cdemqkradr fkqfnvkdlr dyrkhekmpv livvidefqv lfsdnkstka veghlntllk
 481 kgrsygvhlv latqtmrgtd inpsfkaqia nrilpmdae dsssvlgdda aceiqkpegi
 30 541 fnnnggnrky htkmsvpkap ddfksfltki haefnqrnla pidrkiynge tplkmpdtlk

601 anemrlhlhgk kvdyeqkdli vefesneshl lvviqdl nar ialmklfqn vksankelvf
 661 cnkeklrlirs fdagkeygit pvenilsvld tamnpnsalv idnlneakel hdkvgaeklk
 721 sflekaidne qycvifahdf rqiktnyhfd klkellnnhf kqclafrncg enlnaiksdl
 781 pppsklnvll ielskdshte frpfsl

5 SEQ ID NO: 7 Nucleotide coding sequence for VP4 (ADS4) protein

1 ggctataaaa tggcttcgct catttataga caactttctca ctaattcata ttcggtagac
 61 ttgtatgacg aaatagaaca gattggatcg gagaaaactc aaaatgtgac gataaatcca
 121 ggcccttttg cacagactag atatgctcca gttgattggg gacacggaga gattaatgat
 181 tcaactacag tggaaccagt tttagatggc ccgatcaac ccactacatt caaaccaccc
 10 241 aatgattatt ggctgcttat tagctcaaat acagatggag tagtctatga gagtacaaat
 301 aatagtgact tttggacagc agttatcgct gtcgaaccac atgttagtca acaaatagg
 361 caatatgttt tatttggtga gaataagcag tttaacgtag aaaataattc agataaatgg
 421 aaatttttcg aaatgtttta aggtagtagt cagagtgtt tttctaatag acggactcta
 481 acctctaata atagacttgt aggaatgcta aaatatgggt gaagagtatg gacgtttcat
 15 541 ggtgaaacac caagagctac tactgatagt tcgaatactg cggatttaaa taatatatca
 601 attataattc attcagagtt ttatattatt ccaagatccc aagaatctaa gtgtaatgaa
 661 tatattaata atggtttacc accaattcaa aatactagaa atgtagttcc attatctcta
 721 tcatccagat ctattcaata taggagagca caagttaatg aagatattac aatctcaaaa
 781 acttcattat ggaaggaaat gcaatataat agagatatta taataagatt taaatttggt
 20 841 aatagtgtca taaaactagg aggattgggg tataaatggc ctgaaatatc atacaaagca
 901 gcgaattatc aatatagcta ttcacgtgat ggtgaacaag ttactgagca taccacctgt
 961 tcagtaaattg gagtaaataa ttttagctat aatggaggct cgctacctac tgatttcagt
 1021 atttcgagat atgaagttat taaaggaatt cattatgtat atatagacta ctgggatgat
 1081 tcaaaagcat tcagaaatat ggtatatgtt agatcgtag cagcaaattt gaattcagt
 25 1141 aaatgtgtag gtgggagtta tgattttaga ttacctgtag gtgaatggcc tattatgaat
 1201 ggcggtgctg tatcattaca ttttgctgga gtgacattat ctacacagtt cactgatttt
 1261 gtatcattga attcattacg atttagattc agtttaacag tagatgaacc atctttctca
 1321 ataatacgaa cacgtacaat gaatttatac ggattaccag cagctaatac aaacaatgga
 1381 aatgaatact atgaagtgtc aggaagggtt tcaattatct cttagttcc aaccattgat
 30 1441 gattatcaaa ctccaattat gaattcagta acagtaaggc aagattttaga acgccagctt

1501 aatgatttac gagaagaatt taattcattg tcacaagaaa tagctatgtc acaattaatt
 1561 gatttagcat tattaccttt agacatgttt tctatgtttt cggggataaa aagcactatt
 1621 gatctgacca agtcaatggc aactagtgtg atgaaaaaat ttagaaaatc aaaattagcc
 1681 acatcaattt cagaaatgac taattcattg tcagatgcgg cttcgtcagc atcaagaagt
 5 1741 gtttctatta gatcaaattt ttcgacgatt tcaaattggt ctgatgcttc aaaaagtgtg
 1801 ttaaatgtaa ctgactcagt aaatgatgtt tcaacacaaa cttctataat tagtaagaaa
 1861 cttagattaa gagagatgat tactcaaact gaaggaatta gttttgacga tatttcagca
 1921 gctgtactga aaacgaaaat agatatgtcc acacaaattg gaaaaaatac cttacctgat
 1981 gtagttactg aagactctga aaagtttatt ccaaaacgat cgtatcgagt attaaaagat
 10 2041 aatgaagtga tggaaattaa cactgaagga aagttttttg cttataaagt ggatacactc
 2101 aatgagatcc cattcgatat aaataaattc gcggaacttg taacggattc tccagttata
 2161 tcagcgataa tagactttaa gcgattaaaa aatttaaacg ataattatgg aattactcgc
 2221 atagaagcgc ttaatttaat taaatcgaat ccgaatgtac tacgtaattt tattaatcaa
 2281 aataatccaa ttataagaaa tagaattgag cagttaattc tacaatgtaa attgtgagaa
 15 2341 tgtcattgag gatgtgacc

SEQ ID NO: 8 Protein VP4; ADS4

1 masliyrqlt tnsysvdlyd eieqigsekt qnvtinpgpf aqtryapvdw ghgeindstt
 61 vepvldgpyq pttfkppndy wllissntdg vvestnnsd fwtaviavep hvsqtnrqyv
 121 lfgenkqfnv ennsdkwkff emfkqssqsd fsnrtrltsn nrlvgmlkyg grvwtfhget
 20 181 prattddsnt adlnnisiii hsefyiiprs qeskcneyin nglppiqntr nvvplslssr
 241 siqyrraqvn editisktsl wkemqynrdi iirfkfgnsv iklgglgykw seisykaany
 301 qsysrdgeq vtahttcsvn gvnfsyngg slptdfsisr yevikgihyv yidywddska
 361 frnmvyvrsl aanlnsvkcv ggsydfrlpv gewpimngga vslhfagvtl stqftdfvsl
 421 nslrfrfslt vdepsfsiir trtmnlyglp aanpnngney yevsgrfsli slvptiddyq
 25 481 tpimnsvtvr qdlerqlndl reefnslsqe iamsqlidla llpldmfsmf sgikstidlt
 541 ksmatsvmkk frksklatsi semtnslsda assasrsvsi rsnfstisnw sdasksvlnv
 601 tdsyndvstq tsiiskklrl remitqtegi sfddisaavl ktkidmstqi gkntlpdvvt
 661 edsekfipkr syrvlkdnv meintegkff aykvdtlnei pfdinkfael vtdspvisai
 721 idfkrklknl dnygittriea lnliksnpnv lrnfinqnp iirnrieqli lqckl

SEQ ID NO: 9 Nucleotide coding sequence for Ro60 (ADS5) protein

1 attttgcctt ttgttaggt ttcctaaaga caaaaaaaaaa tggaggaatc tgtaaaccac
 61 atgcagccac tgaatgagaa gcagatagcc aattctcagg atggatatgt atggcaagtc
 121 actgacatga atcgactaca ccggttctta tgtttcgggt ctgaagggtg gacttattat
 5 181 atcaaagaac agaagttggg ccttgaaaat gctgaagctt taattagatt gattgaagat
 241 ggcagaggat gtgaagtgat acaagaaata aagtcattta gtcaagaagg cagaaccaca
 301 aagcaagagc ctatgctctt tgcacttgcc atttgttccc agtgcctcga cataagcaca
 361 aaacaagcag catttaaagc tgtttctgaa gtttgcgcga ttcctaccca tctctttact
 421 tttatccagt ttaagaaaga tctgaaggaa agcatgaaat gtggcatgtg gggtcgtgcc
 10 481 ctccggaagg ctatagcgga ctggtacaat gagaaagggt gcatggccct tgctctggca
 541 gttacaaaat ataaacagag aaatggctgg tctcacaag atctattaag attgtcacat
 601 cttaaaccct ccagtgaagg acttgcaatt gtgaccaa atattacaaa gggctggaaa
 661 gaagttcatg aattgtataa agaaaaagca ctctctgtgg agactgaaaa attattaaag
 721 tatctggagg ctgtagagaa agtgaagcgc aaaaaagatg agctagaagt cattcatcta
 15 781 atagaagaac atagattagt tagagaacat cttttaacaa atcacttaaa gtctaagag
 841 gtatggaagg ctttgttaca agaatgccg ctactgcat tactaaggaa tctaggaaag
 901 atgactgcta attcagtact tgaaccagga aattcagaag tatctttagt atgtgaaaaa
 961 ctgtgtaatg aaaaactatt aaaaaaggct cgtatacatc catttcatat tttgatcgca
 1021 ttagaaactt acaagacagg tcatggtctc agagggaaac tgaagtggcg cctgatgaa
 20 1081 gaaattttga aagcattgga tgctgctttt tataaaacat ttaagacagt tgaaccaact
 1141 ggaaaacggt tcttactagc tgttgatgac agtgcttcta tgaaccaaag agttttgggt
 1201 agtataactca acgctagtac agttgctgca gcaatgtgca tggttgtcac acgaacagaa
 1261 aaagattctt atgtagttgc tttttccgat gaaatggtag catgtccagt gactacagat
 1321 atgaccttac aacagggttt aatggctatg agtcagatcc cagcaggtgg aactgattgc
 25 1381 tctcttccaa tgatctgggc tcagaagaca aacacacctg ctgatgtctt cattgtattc
 1441 actgataatg agacctttgc tggagggtgc catcctgcta ttgctctgag ggagtatcga
 1501 aagaaaatgg atattccagc taaattgatt gtttgtggaa tgacatcaaa tggtttcacc
 1561 attgcagacc cagatgatag aggcattgtg gatatgtgcg gctttgatac tggagctctg
 1621 gatgtaattc gaaatttcac attagatatg atttaacat aagcagcagc acgatccaga
 30 1681 gatccattgc catcagtgat ctactaaaa aatatacagc tacttcccag ctaatctcca

1741 cccaatgaat gätgatggta tagtatgtgc ataatggaaa gttaccttac tg

SEQ ID NO: 10 Protein Ro60; (ADS5)

1 meesvnmqpp lnekqiansq dgyvwqvtdm nrlhrflcfg seggtyyike qklglенаа
61 lirliedgrg cevigeiksf sqegrttkqe pmlfalaics qcsdistkqa afkavsevcr
5 121 ipthlftfiq fkkdlkesmk cgmwgralrk aiadwynekg gmalalavtk ykqrngwshk
181 dllrlshlkp sseglaivtk yitkgwkevh elykekalsv etekllyle avekvkrtd
241 elevihliee hrlvrehlit nhlkskevkw allqemplta llrnlgkmta nsvlepgnse
301 vslvceklcn eklkkarih pphilialet yktghglrgk lkwrpdeeil kaldaafykt
361 fktveptgkr fillavdsas mnqrvlgsil nastvaaamc mvvtrtekds yvafsdemv
10 421 pcptvtdmtl qqvlmamsqi paggtcdslp miwaqktntp advfivftdn etfaggvghpa
481 ialreyrkkm dipaklivcg mtsngftiad pddrgmldmc gfdtgaldvi rnftldmi

CLAIMS

1. A polypeptide, which polypeptide:

(i) has the amino acid sequence as recited in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10 ;

5 (ii) is a fragment thereof having adhesion molecule activity or having an antigenic determinant in common with the polypeptide of (i); or

(iii) is a functional equivalent of (i) or (ii).

2. A polypeptide which is a functional equivalent according to claim 1(iii), is homologous to the amino acid sequence as recited in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10 , and has adhesion molecule activity.

3. A fragment or functional equivalent according to claim 1 or claim 2, which has greater than 50% sequence identity with the amino acid sequence recited in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10 , or with active fragments thereof, preferably greater than 60%, 70%, 80%, 90%, 95%, 98% or 99% sequence identity.

4. A functional equivalent according to any one of claims 1-3, which exhibits significant structural homology with a polypeptide having the amino acid sequence given in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10 .

5. A fragment as recited in claim 1 or claim 3 having an antigenic determinant in common with the polypeptide of claim 1(i), which consists of 7 or more (for example, 8, 10, 12, 14, 16, 18, 20 or more) amino acid residues from the sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or SEQ ID NO: 10 .

6. A purified nucleic acid molecule which encodes a polypeptide according to any one of the preceding claims.

25 7. A purified nucleic acid molecule according to claim 6, which has the nucleic acid sequence as recited in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 or SEQ ID NO: 9 or is a redundant equivalent or fragment thereof.

8. A purified nucleic acid molecule which hybridizes under high stringency conditions

with a nucleic acid molecule according to claim 6 or claim 7.

9.A vector comprising a nucleic acid molecule as recited in any one of claims 6-8.

10.A host cell transformed with a vector according to claim 9.

11.A ligand which binds specifically to, and which preferably inhibits the adhesion
5 molecule activity of, a polypeptide according to any one of claims 1-5.

12.A ligand according to claim 11, which is an antibody.

13.A compound that either increases or decreases the level of expression or activity of a
polypeptide according to any one of claims 1-5.

14.A compound according to claim 13 that binds to a polypeptide according to any one of
10 claims 1-5 without inducing any of the biological effects of the polypeptide.

15.A compound according to claim 14, which is a natural or modified substrate, ligand,
enzyme, receptor or structural or functional mimetic.

16.A polypeptide according to any one of claim 1-5, a nucleic acid molecule according to
any one of claims 6-8, a vector according to claim 9, a ligand according to claim 11
15 or 12, or a compound according to any one of claims 13-15, for use in therapy or
diagnosis of disease.

17.A method of diagnosing a disease in a patient, comprising assessing the level of
expression of a natural gene encoding a polypeptide according to any one of claim 1-
5, or assessing the activity of a polypeptide according to any one of claim 1-5, in
20 tissue from said patient and comparing said level of expression or activity to a control
level, wherein a level that is different to said control level is indicative of disease.

18.A method according to claim 17 that is carried out in vitro.

19.A method according to claim 17 or claim 18, which comprises the steps of: (a)
contacting a ligand according to claim 11 or claim 12 with a biological sample under
25 conditions suitable for the formation of a ligand-polypeptide complex; and (b)
detecting said complex.

20.A method according to claim 17 or claim 18, comprising the steps of:

- a) contacting a sample of tissue from the patient with a nucleic acid probe under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 6-8 and the probe;
- b) contacting a control sample with said probe under the same conditions used in step a);
- 5 and
- c) detecting the presence of hybrid complexes in said samples; wherein detection of levels of the hybrid complex in the patient sample that differ from levels of the hybrid complex in the control sample is indicative of disease.

21.A method according to claim 17 or claim 18, comprising:

- 10 a) contacting a sample of nucleic acid from tissue of the patient with a nucleic acid primer under stringent conditions that allow the formation of a hybrid complex between a nucleic acid molecule according to any one of claims 6-8 and the primer;
- b) contacting a control sample with said primer under the same conditions used in step a);
- and
- 15 c) amplifying the sampled nucleic acid; and
- d) detecting the level of amplified nucleic acid from both patient and control samples; wherein detection of levels of the amplified nucleic acid in the patient sample that differ significantly from levels of the amplified nucleic acid in the control sample is indicative of disease.

20 22.A method according to claim 17 or claim 18 comprising:

- a) obtaining a tissue sample from a patient being tested for disease;
- b) isolating a nucleic acid molecule according to any one of claims 6-8 from said tissue sample; and
- c) diagnosing the patient for disease by detecting the presence of a mutation which is
- 25 associated with disease in the nucleic acid molecule as an indication of the disease.

23.The method of claim 22, further comprising amplifying the nucleic acid molecule to form an amplified product and detecting the presence or absence of a mutation in the amplified product.

24. The method of either claim 22 or 23, wherein the presence or absence of the mutation in the patient is detected by contacting said nucleic acid molecule with a nucleic acid probe that hybridises to said nucleic acid molecule under stringent conditions to form a hybrid double-stranded molecule, the hybrid double-stranded molecule having an unhybridised portion of the nucleic acid probe strand at any portion corresponding to a mutation associated with disease; and detecting the presence or absence of an unhybridised portion of the probe strand as an indication of the presence or absence of a disease-associated mutation.
25. A method according to any one of claims 17-24, wherein said disease is cardiovascular disease, cancer, asthma, chronic obstructive pulmonary disease (COPD), inflammatory disease or a bacterial infection.
26. Use of a polypeptide according to any one of claims 1-5 as an adhesion molecule.
27. A pharmaceutical composition comprising a polypeptide according to any one of claim 1-5, a nucleic acid molecule according to any one of claims 6-8, a vector according to claim 9, a ligand according to claim 11 or 12, or a compound according to any one of claims 13-15.
28. A vaccine composition comprising a polypeptide according to any one of claims 1-5 or a nucleic acid molecule according to any one of claims 6-8.
29. A polypeptide according to any one of claims 1-5, a nucleic acid molecule according to any one of claims 6-8, a vector according to claim 9, a ligand according to claim 11 or 12, a compound according to any one of claims 13-15, or a pharmaceutical composition according to claim 21, for use in the manufacture of a medicament for the treatment of cardiovascular disease, cancer, asthma, chronic obstructive pulmonary disease (COPD), inflammatory disease or a bacterial infection.
30. A method of treating a disease in a patient, comprising administering to the patient a polypeptide according to any one of claims 1-5, a nucleic acid molecule according to any one of claims 6-8, a vector according to claim 9, a ligand according to claim 11 or 12, a compound according to any one of claims 13-15, or a pharmaceutical composition according to claim 21.

31. A method according to claim 30, wherein, for diseases in which the expression of the natural gene or the activity of the polypeptide is lower in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist.

32. A method according to claim 30, wherein, for diseases in which the expression of the natural gene or activity of the polypeptide is higher in a diseased patient when compared to the level of expression or activity in a healthy patient, the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an antagonist.

33. A method of monitoring the therapeutic treatment of disease in a patient, comprising monitoring over a period of time the level of expression or activity of a polypeptide according to any one of claims 1-5, or the level of expression of a nucleic acid molecule according to any one of claims 6-8 in tissue from said patient, wherein altering said level of expression or activity over the period of time towards a control level is indicative of regression of said disease.

34. A method for the identification of a compound that is effective in the treatment and/or diagnosis of disease, comprising contacting a polypeptide according to any one of claims 1-5, or a nucleic acid molecule according to any one of claims 6-8 with one or more compounds suspected of possessing binding affinity for said polypeptide or nucleic acid molecule, and selecting a compound that binds specifically to said nucleic acid molecule or polypeptide.

35. A kit useful for diagnosing disease comprising a first container containing a nucleic acid probe that hybridises under stringent conditions with a nucleic acid molecule according to any one of claims 6-8; a second container containing primers useful for amplifying said nucleic acid molecule; and instructions for using the probe and primers for facilitating the diagnosis of disease.

36. The kit of claim 35, further comprising a third container holding an agent for digesting unhybridised RNA.

37. A kit comprising an array of nucleic acid molecules, at least one of which is a nucleic acid molecule according to any one of claims 6-8.

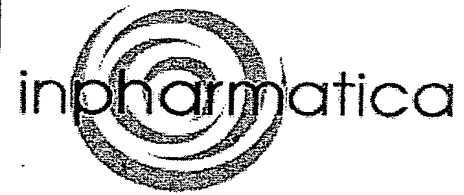
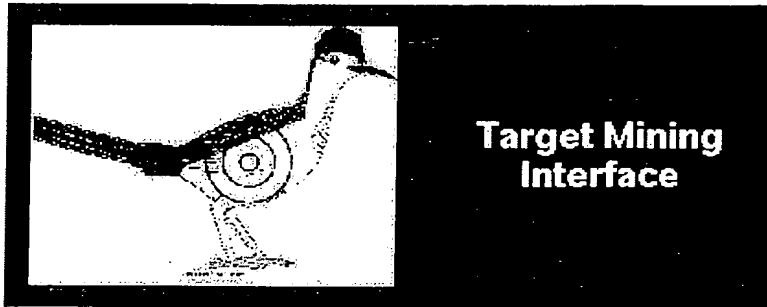
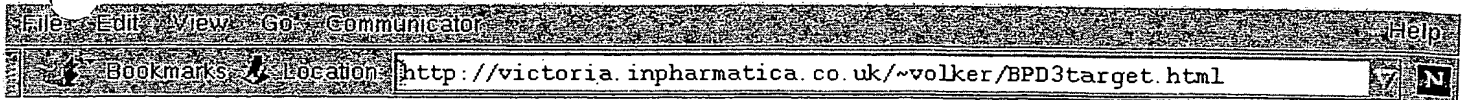
38. A kit comprising one or more antibodies that bind to a polypeptide as recited in any one of claims 1-5; and a reagent useful for the detection of a binding reaction between
5 said antibody and said polypeptide.

39. A transgenic or knockout non-human animal that has been transformed to express higher, lower or absent levels of a polypeptide according to any one of claims 1-5.

40. A method for screening for a compound effective to treat disease, by contacting a non-human transgenic animal according to claim 39 with a candidate compound and
10 determining the effect of the compound on the disease of the animal.

THIS PAGE BLANK (USPTO)

Figure 1



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:

☐ Homo sapiens ☐ Rattus norvegicus (Rat) ☐ Mus musculus (Mouse) ☐ Danio rerio (Zebrafish)

THIS PAGE BLANK (USPTO)

2) 84 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st Iter. (PSI)	Best Iter. (PSI)	Res E-va (PS)
<input type="checkbox"/>	AAA59544.1 drill through Top50BlastHits Red Seq View	AAA59544.1	Not given	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E-
<input checked="" type="checkbox"/>	AAB24821.1 drill through Top50BlastHits Red Seq View	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	33.7%, 33% unmaskedSW	4-183, 4-183	150-336, 150-336	449	100% unmaskedGT	1	2	3E-
<input type="checkbox"/>	Q99715 drill through Top50BlastHits Red Seq View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.8%, 26% unmaskedSW	4-181, 2-174	140-318, 2321-2495	440	100% unmaskedGT	1	3	2E-
<input type="checkbox"/>	P20701 drill through Top50BlastHits Red Seq View	P20701	LEUKOCYTE ADHESION GLYCOPROTEIN LFA-1 ALPHA CHAIN PRECURSOR (LEUKOCYTE FUNCTION ASSOCIATED MOLECULE 1, ALPHA CHAIN)(CD11A) (INTEGRIN ALPHA-L).	Homo sapiens (Human).	PRI	100%, 100% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E-
<input type="checkbox"/>	AAC31672.1 drill through Top50BlastHits Red Seq View	AAC31672.1	leukocyte function-associated molecule-1 alpha subunit	Homo sapiens	PRI	99.5%, 99% unmaskedSW	1-183, 1-183	153-335, 153-335	423	100% unmaskedGT	1	1	1E-
<input type="checkbox"/>	CAA72402.1 drill through Top50BlastHits Red Seq View	CAA72402.1	collagen type XIV	Homo sapiens	PRI	29.1%, 29% unmaskedSW	2-180, 2-180	5-185, 5-185	422	100% unmaskedGT	1	2	2E-
<input type="checkbox"/>	AAB38702.1 drill through Top50BlastHits Red Seq View	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	31.7%, 27% unmaskedSW	4-183, 2-182	275-455, 39-223	413	100% unmaskedGT	1	2	2E-
<input type="checkbox"/>	CAB70853.1 drill through Top50BlastHits Red Seq View	CAB70853.1	hypothetical protein	Homo sapiens	PRI	28%, 28% unmaskedSW	1-180, 1-183	437-620, 437-624	406	100% unmaskedGT	1	2	1E-
<input type="checkbox"/>	CAA27972.1 drill through Top50BlastHits Red Seq View	CAA27972.1	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	1497-1673, 1689-1873	405	100% unmaskedGT	2	3	1E-
<input type="checkbox"/>	AAB59512.1 drill through Top50BlastHits Red Seq View	AAB59512.1	Not given	Homo sapiens	PRI	20.5%, 20% unmaskedSW	3-183, 2-181	758-934, 950-1134	405	100% unmaskedGT	2	3	1E-
<input type="checkbox"/>	CAA07569.1 drill through Top50BlastHits Red Seq View	CAA07569.1	matrilin-4	Homo sapiens	PRI	28.1%, 25% unmaskedSW	1-183, 1-183	342-528, 31-217	403	100% unmaskedGT	1	2	2E-

THIS PAGE BLANK (USPTO)

Figure 2B

1) 509 hits identified by Genome Threader only:

Adk2list	Redundant sequence display	BPD link	WWW link	Title	Organism	Div.	%ID	Query rgn.	Target rgn.	Aln. score	Conf.
<input type="checkbox"/>	Red. Seq. View	AA71133.1 drill through Top50BlastHits	AA71133.1	PRD2769	Homo sapiens	PRI	<u>13.8%</u> unmaskedSW	109-179	1-80	122	<u>100%</u> unmaskedGI
<input type="checkbox"/>	Red. Seq. View	CAB52192.1 drill through Top50BlastHits	CAB52192.1	G7c protein	Homo sapiens	PRI	<u>9.6%</u> unmaskedSW	7-117	20-124	82	<u>99.18%</u> unmaskedGI
<input type="checkbox"/>	Red. Seq. View	CAA82910.1 drill through Top50BlastHits	CAA82910.1	basic transcription factor 2, 44 kD subunit	Homo sapiens	PRI	<u>11%</u> unmaskedSW	5-164	61-225	75	<u>99.18%</u> unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAD21820.1 drill through Top50BlastHits	AAD21820.1	NG37	Homo sapiens	PRI	<u>9.6%</u> unmaskedSW	7-117	318-422	82	<u>98.86%</u> unmaskedGI
<input type="checkbox"/>	Red. Seq. View	BAA20761.1 drill through Top50BlastHits	BAA20761.1	Not given	Homo sapiens	PRI	<u>16.4%</u> unmaskedSW	5-114	1836-1950	78	<u>98.52%</u> unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAA36154.1 drill through Top50BlastHits	AAA36154.1	Not given	Homo sapiens	PRI	<u>13.2%</u> unmaskedSW	5-112	10-137	79	<u>98.16%</u> unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAF03046.1 drill through Top50BlastHits	AAF03046.1	candidate tumor suppressor protein DICE1	Homo sapiens	PRI	<u>13.8%</u> unmaskedSW	5-113	4-131	79	<u>97.07%</u> unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAC74854.1 drill through Top50BlastHits	AAC74854.1	orf, hypothetical protein	Escherichia coli	BCT	<u>16.7%</u> unmaskedSW	5-99	250-335	78	<u>95.09%</u> unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAB60942.1 drill through Top50BlastHits	AAB60942.1	breast cancer suppressor candidate 1	Homo sapiens	PRI	<u>16.2%</u> unmaskedSW	63-156	2-90	80	<u>93.33%</u> unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AA67537.1 drill through Top50BlastHits	AA67537.1	glycoprotein IIIa	Homo sapiens	PRI	<u>18.9%</u> unmaskedSW	3-112	111-251	72	<u>92.08%</u> unmaskedGI reverse Hit
<input type="checkbox"/>	Red. Seq. View	AA52589.1 drill through Top50BlastHits	AA52589.1	Not given	Homo sapiens	PRI	<u>19.6%</u> unmaskedSW	3-112	138-277	74	<u>91.34%</u> unmaskedGI reverse Hit
<input type="checkbox"/>	Red. Seq. View	AAB71380.1 drill through Top50BlastHits	AAB71380.1	platelet membrane glycoprotein IIIa beta subunit	Homo sapiens	PRI	<u>19.6%</u> unmaskedSW	3-112	138-277	74	<u>91.34%</u> unmaskedGI reverse Hit
<input type="checkbox"/>	Red. Seq. View	AAA35927.1 drill through Top50BlastHits	AAA35927.1	Not given	Homo sapiens	PRI	<u>19.6%</u> unmaskedSW	3-112	138-277	74	<u>91.34%</u> unmaskedGI reverse Hit

THIS PAGE BLANK (USPTO)

630 out of these 645 PSI-BLAST matches were identified using 'positive iterations'.

[illegible]

15 out of these 645 PSI-BLAST matches were identified using 'negative iterations':

BAA91233.1,CMAI2066.1,CAE50193.1,CAA17374.1,AES86281.1,ABD0774.1,AA93386.1,AA93386.1,AA93386.1,AA93386.1,ARC16991.1,EAA6551.1,ABA61649.1,ANC51323.1,ANC52846.1,Q01443.

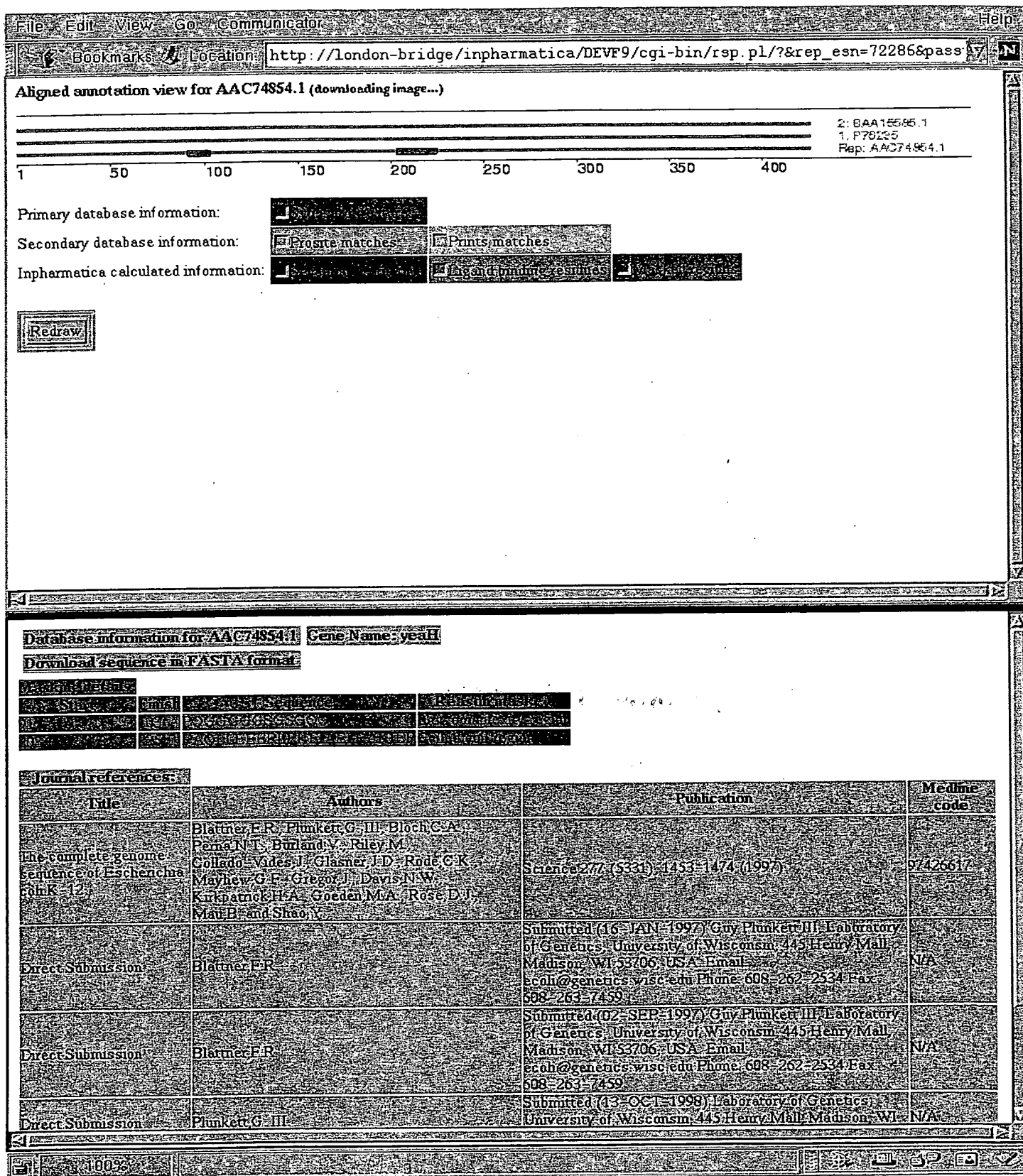
A2) Genome Threader Matches:

2729 matches found by Genome Threading:

[illegible]

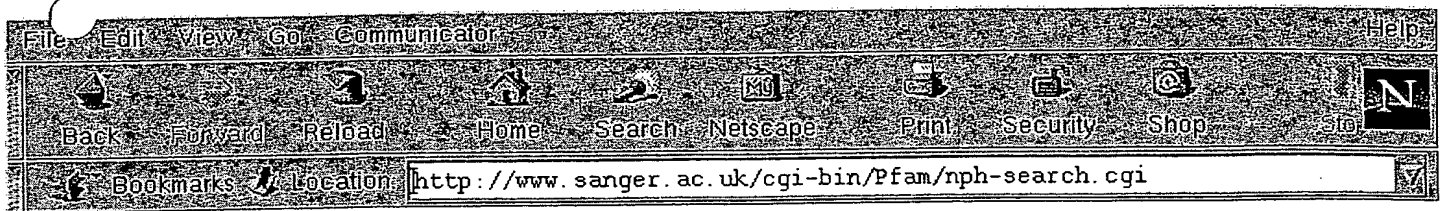
THIS PAGE BLANK (USPTO)

Figure 3



THIS PAGE BLANK (USPTO)

Figure 4



Results for gi|1788084|gb|AAC74854.1|

There were no matches to Pfam-A (including borderline matches) for gi|1788084|gb|AAC74854.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B 39416	233	423	3.7e-103	Align

[427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families:

[Hypertext linked to swisspfam](#)

Query gi|1788084|gb|AAC74854.1|233-423 matching Pfam-B 39416

```

YEAM_ECOLI 233 DLRYKNYEKRPDPSSQAVMFCCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
gi|1788084|gb|AAC74854.1| 233 DLRYKNYEKRPDPSSQAVMFCCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
YEAM_ECOLI 283 TYKNVEVVYIRHGTQAKEVDEMEFFYSQETGGTIVSSALKLMDEVVKERY 332
gi|1788084|gb|AAC74854.1| 283 TYKNVEVVYIRHGTQAKEVDEMEFFYSQETGGTIVSSALKLMDEVVKERY 332
YEAM_ECOLI 333 NPAQWNIYAAQASDGDNDWADD SPLCHEILAKKLLPVVRYYSYIEITRRAM 382
gi|1788084|gb|AAC74854.1| 333 NPAQWNIYAAQASDGDNDWADD SPLCHEILAKKLLPVVRYYSYIEITRRAM 382
YEAM_ECOLI 383 QTLWREYEHQLSTFDNFAMQHIRDQDDIYPVFRELPHKQNA 423
gi|1788084|gb|AAC74854.1| 383 QTLWREYEHQLSTFDNFAMQHIRDQDDIYPVFRELPHKQNA 423

```

[Align to family](#)

If you think there is anything wrong with this script, please contact [Pfam](#)



THIS PAGE BLANK (USPTO)

que 5

File Edit View Go Communicator		Help
Bookmarks Location: http://www.ncbi.nlm.nih.gov/80/entrez/query.fcgi?cmd=Retrieve&db=		
LOCUS	AAC74854 427 aa BCT 01-DEC-2000	
DEFINITION	orf, hypothetical protein [Escherichia coli K12].	
ACCESSION	AAC74854	
PID	g1788084	
VERSION	AAC74854.1 GI:1788084	
DBSOURCE	locus AE000273 accession AE000273.1	
KEYWORDS		
SOURCE	Escherichia coli K12.	
ORGANISM	Escherichia coli K12 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.	
REFERENCE	1 (residues 1 to 427)	
AUTHORS	Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.	
TITLE	The complete genome sequence of Escherichia coli K-12	
JOURNAL	Science 277 (5331), 1453-1474 (1997)	
MEDLINE	97426617	
PUBMED	9278503	
REFERENCE	2 (residues 1 to 427)	
AUTHORS	Blattner, F.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	
REFERENCE	3 (residues 1 to 427)	
AUTHORS	Blattner, F.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	
REFERENCE	4 (residues 1 to 427)	
AUTHORS	Plunkett, G. III.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA	
COMMENT	This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@ambr.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible at (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names. Method: conceptual translation.	
FEATURES	Location/Qualifiers	
source	1..427 /organism="Escherichia coli K12" /strain="K12" /sub_strain="MG1655" /db_xref="taxon:83333"	
Protein	1..427 /function="orf, Unknown" /product="orf, hypothetical protein"	
CDS	1..427 /gene="peaH" /coded_by="1788078:6385..7668" /transl_table=11 /note="o427; This 427 aa ORF is 28 pct identical (43 gaps) to 327 residues of an approx. 312 aa protein YZDC_BAC5U SW: P45742"	
ORIGIN	1 mtwfidrrln gknksmvnrq rllrrypkqi kqsiseaink rsvtdvdsge svsiptedis 61 epmfhgqrgg lhrvhpgnd hfvqndrier pggggggsgs gggqasqdge ggdefvfiqs 121 kdeyldilfe dlalpnlkqm qqrqlteykt hragytagv panisvvrsl qnslarrtam 181 tagkrrelha leenlaiisn sepaqlleee rlrkeiaelr akiervpfid tfdlryknye	

THIS PAGE BLANK (USPTO)

Figure 6A

File Edit Search Type Help

Accession Code Query

inpharmatica

TOOLS

←

→

?

⌂

⌂

X

Sequences

? ? ? ? ?

? Acc y123

Ligands

? ADP NAD FAD

? Phe Ala Glu

Select a database and enter a code:

Database	
GenBank	
Accession	<input checked="" type="radio"/>
Accession Version	<input type="radio"/>
GI-NID	<input type="radio"/>
protein-id	<input type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

AAC74854.1

Submit Reset

THIS PAGE BLANK (USPTO)

Aligned Sequence Display

Query details:

AAC74854.1: orf, hypothetical protein



informatics

<div> <div>Tools</div> <div> <div> <div>←</div> <div>→</div> <div>?</div> <div>📄</div> <div>📄</div> <div>✕</div> </div> <div>Sequences</div> <div> <div>🔍</div> <div>🔍</div> <div>🔍</div> <div>🔍</div> </div> <div>Ligands</div> <div> <div>🔍</div> <div>🔍</div> <div>🔍</div> </div> </div> </div>										
<div> <div>Total selected: 0</div> <div>View alignment</div> </div>										
<div> <div> <div>Page 1 of 2</div> <div> <div>Total hits: 36</div> <div>Selected: 0</div> </div> </div> <div> <div>Select all</div> <div>Deselect all</div> </div> </div>										
<div> <div>Propharmatica Genome Threader results</div> </div>										
Cluster	Detail	Code	Title	Organism	% ID	Query (gn)	Target (gn)	Align. score	Method	Confidence
1	→	1LFA:A	CD11A I-DOMAIN WITH BOUND MN++	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal (95%)
	▶	1ZOO:A	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal (95%)
	▶	1ZOO:B	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal (95%)
	▶	1ZOP:B	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal (95%)
	▶	1ZOP:A	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	78	Local	Marginal (95%)
2	▶	1BHO:1	MAC-1 I-DOMAIN MAGNESIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low (94%)
	▶	11DN:1	MAC-1 I-DOMAIN METAL FREE	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low (94%)
	▶	1BHQ:2	MAC-1 I-DOMAIN CADMIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low (94%)
	▶	1BHQ:1	MAC-1 I-DOMAIN CADMIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low (94%)
	▶	1BHQ:2	MAC-1 I-DOMAIN MAGNESIUM COMPLEX	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low (94%)
3	▶	11DN:2	MAC-1 I-DOMAIN METAL FREE	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	76	Local	Low (94%)
	▶	1ATZ:B	HUMAN VON WILLEBRAND FACTOR A3 DO...	<i>H.sapiens</i>	15.1	244 - 352	1 - 126	70	Local	Low (94%)
	▶	1ATZ:A	HUMAN VON WILLEBRAND FACTOR A3 DO...	<i>H.sapiens</i>	15.3	250 - 335	5 - 101	72	Local	Low (93%)
	▶	1DGO:A	NMR SOLUTION STRUCTURE OF THE INSE...	<i>H.sapiens</i>	16.3	250 - 331	9 - 99	75	Local	Low (93%)
	▶	1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BO...	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	75	Local	Low (92%)
4	▶	1JEM:1	I-DOMAIN FROM INTEGRIN CR3, MN2+ BO...	<i>H.sapiens</i>	15.1	250 - 365	4 - 128	75	Local	Low (89%)
	▶	1ZON	CD11A I-DOMAIN WITHOUT BOUND CATI...	<i>H.sapiens</i>	16.7	250 - 335	5 - 99	72	Local	Low (89%)
	▶	1LEA:B	CD11A I-DOMAIN WITH BOUND MN++	<i>H.sapiens</i>	16.0	250 - 333	5 - 97	66	Local	Low (80%)
	▶	1OAK:A	CRYSTAL STRUCTURE OF THE VON WILLE...	<i>H.sapiens</i>	20.3	251 - 301	1 - 59	72	Local	Low (72%)
	▶	1AUQ	A1 DOMAIN OF VON WILLEBRAND FACTOR	<i>H.sapiens</i>	21.2	250 - 301	18 - 68	69	Local	Low (66%)
5	▶	1A03:A	A3 DOMAIN OF VON WILLEBRAND FACTOR	<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low (53%)
	▶	1A03:B	A3 DOMAIN OF VON WILLEBRAND FACTOR	<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low (53%)
	▶	1A0X:B	I DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>H.sapiens</i>	13.1	242 - 335	1 - 105	53	Local	Low (21%)
	▶	1RYP:D	CRYSTAL STRUCTURE OF THE 20S PROTE...	<i>S.cerevisiae</i>	13.1	246 - 305	166 - 226	58	Local	Low (21%)
	▶	1RYP:R	CRYSTAL STRUCTURE OF THE 20S PROTE...	<i>S.cerevisiae</i>	13.1	246 - 305	166 - 226	58	Local	Low (21%)
6	▶	1CK4:B	CRYSTAL STRUCTURE OF RAT A1B1 INTEG...	<i>R.nonvegicus</i>	10.5	250 - 331	5 - 96	55	Local	Low (20%)
	▶	1A03:A	A3 DOMAIN OF VON WILLEBRAND FACTOR	<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low (53%)
	▶	1A03:B	A3 DOMAIN OF VON WILLEBRAND FACTOR	<i>H.sapiens</i>	15.2	245 - 331	2 - 99	63	Local	Low (53%)
	▶	1A0X:B	I DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>H.sapiens</i>	13.1	242 - 335				

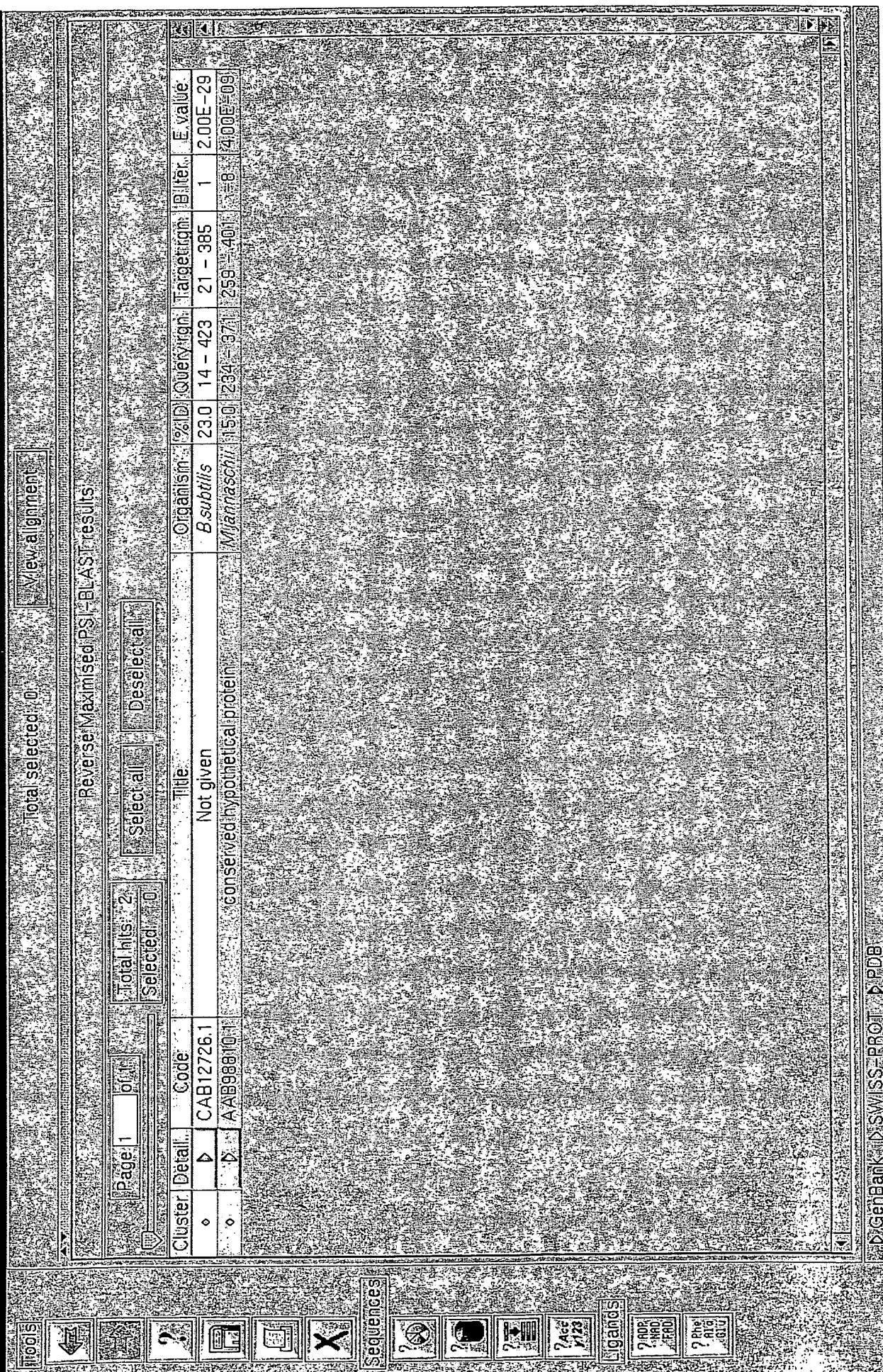
THIS PAGE BLANK (USPTO)

Aligned Sequence Display

Query details:

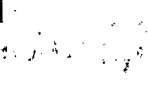
AAC74854.1:orf, hypothetical protein

Figure 6C



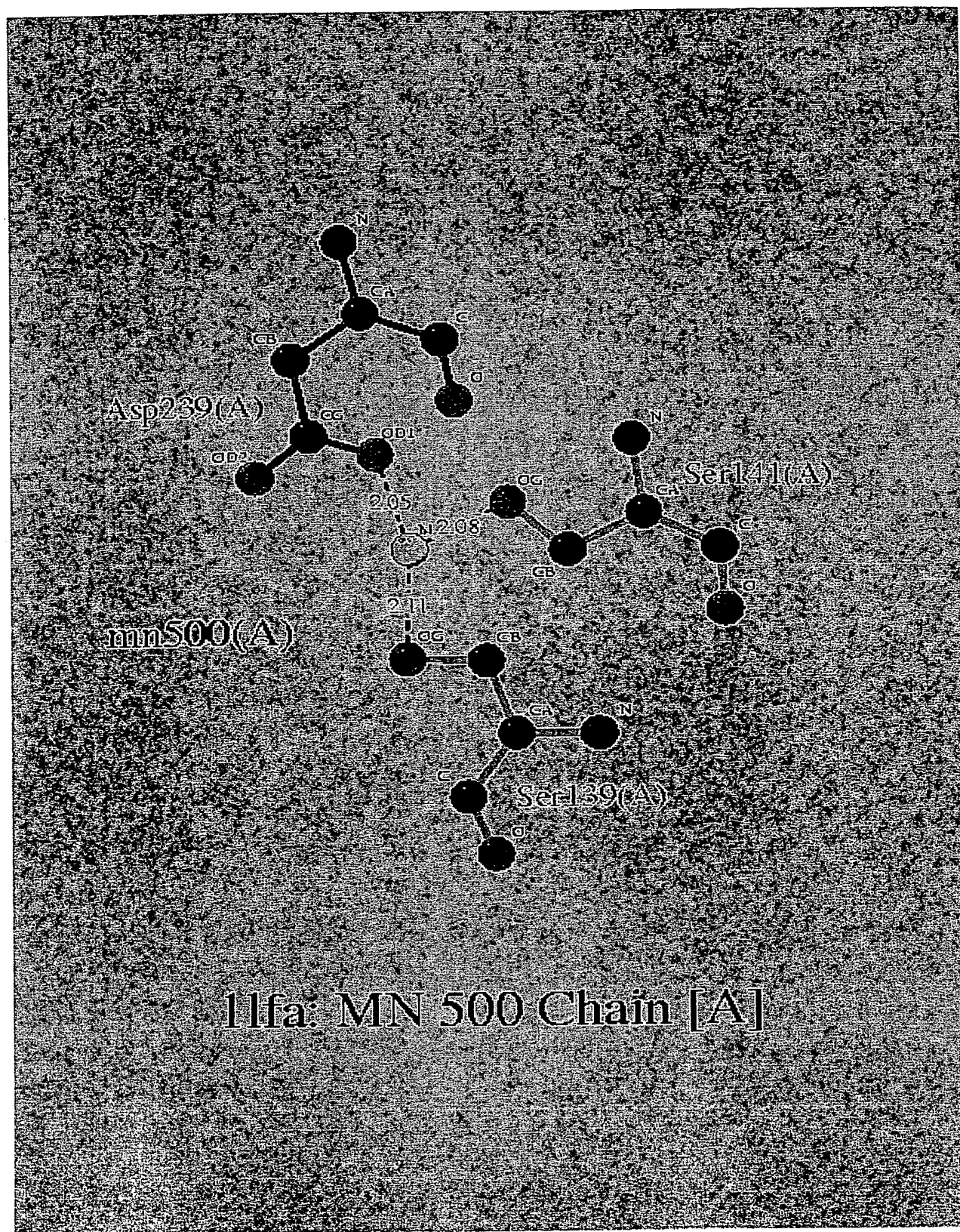
THIS PAGE BLANK (USPTO)

Figure 7



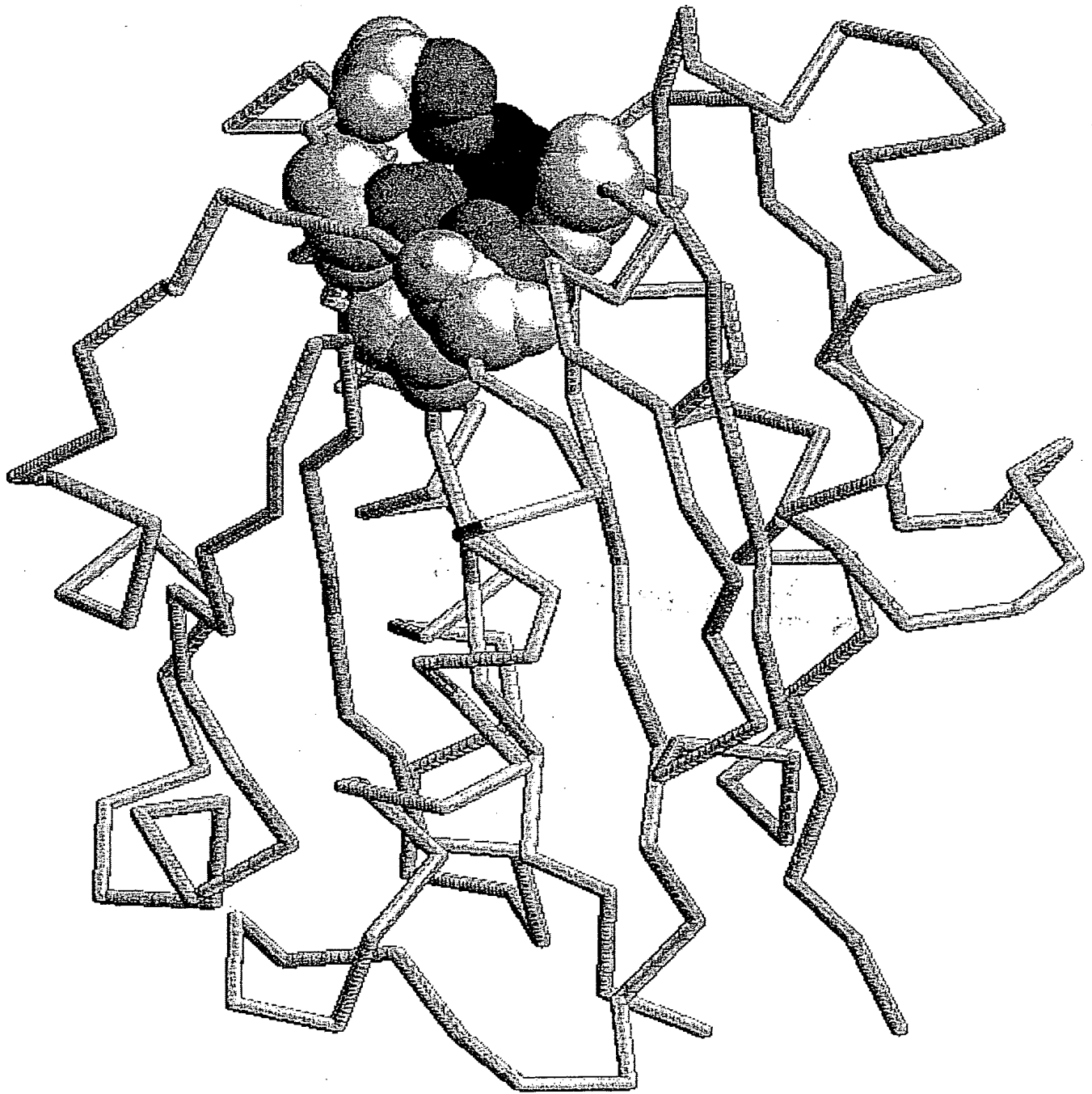
THIS PAGE BLANK (USPTO)

Figure 8a



THIS PAGE BLANK (USPTO)

● Figure 8B
○

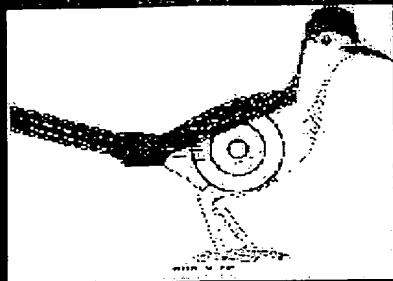


THIS PAGE BLANK (USPTO)


● Figure 9

File Edit View Go Communicator Help

Bookmarks Location: <http://victoria.inpharmatica.co.uk/~volker/BPD3target.html>



Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

THIS PAGE BLANK (USPTO)

Figure 10A

[File](#)
[Edit](#)
[View](#)
[Go](#)
[Communicator](#)

[Back](#)
[Forward](#)
[Reload](#)
[Home](#)
[Search](#)
[Netscape](#)
[Print](#)
[Security](#)
[Shop](#)

[Bookmarks](#)
[Location](#)
[http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl#Mine2](#)

2) 82 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in **maroon**!

Add list	BPD links	WWW link	Title	Organism	Div.	%ID (GI/PSI)	Query rgn. (GI/PSI)	Target rgn. (GI/PSI)	Aln. score (GI)	Conf. (GI)	1st iter. (PSI)	Best iter. (PSI)	Best E-value (PSI)
<input type="checkbox"/>	P56199 drill through Top50BlastHits Red Seq View	P56199	INTEGRIN ALPHA-1 (LAMININ AND COLLAGEN RECEPTOR)(VLA-1) (CD49A).	Homo sapiens (Human).	PRI	51.8%, 51% unmaskedSW	2-200, 2-200	139-337, 139-337	487	100% unmaskedGI	1	2	2E-73
<input type="checkbox"/>	P17301 drill through Top50BlastHits Red Seq View	P17301	PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GPIA) (COLLAGEN RECEPTOR)(INTEGRIN ALPHA-2)(VLA-2 ALPHA CHAIN) (CD49B).	Homo sapiens (Human).	PRI	99.5%, 100% unmaskedSW	2-200, 2-200	189-367, 189-367	471	100% unmaskedGI	1	1	1E-111
<input type="checkbox"/>	AAFD1258.1 drill through Top50BlastHits Red Seq View	AAFD1258.1	integrin alpha-11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmaskedSW	2-200, 2-200	159-355, 159-355	469	100% unmaskedGI	1	2	4E-68
<input type="checkbox"/>	AAD51919.2 drill through Top50BlastHits Red Seq View	AAD51919.2	integrin alpha 11 subunit precursor	Homo sapiens	PRI	45.2%, 45% unmaskedSW	2-200, 2-200	158-355, 159-355	469	100% unmaskedGI	1	2	4E-68
<input type="checkbox"/>	Q99715 drill through Top50BlastHits Red Seq View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.1%, 31% unmaskedSW	2-196, 7-189	133-318, 2323-2495	437	100% unmaskedGI	1	2	2E-42
<input type="checkbox"/>	AAAS9491.1 drill through Top50BlastHits Red Seq View	AAAS9491.1	Not given	Homo sapiens	PRI	27.1%, 27% unmaskedSW	2-200, 2-200	144-338, 144-338	435	100% unmaskedGI	1	2	1E-60
<input type="checkbox"/>	AAAS9544.1 drill through Top50BlastHits Red Seq View	AAAS9544.1	Not given	Homo sapiens	PRI	27.1%, 27% unmaskedSW	2-200, 2-200	144-338, 144-338	435	100% unmaskedGI	1	2	1E-60
<input type="checkbox"/>	CAB71222.1 drill through Top50BlastHits Red Seq View	CAB71222.1	α2(38D)15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	26.9%, 26% unmaskedSW	5-195, 156-201	293-472, 1581-1828	424	100% unmaskedGI	2	3	5E-6
<input type="checkbox"/>	P21941 drill through Top50BlastHits Red Seq View	P21941	CARTILAGE MATRIX PROTEIN PRECURSOR (MATRILIN-1).	Homo sapiens (Human).	PRI	26.8%, 26% unmaskedSW	3-198, 3-198	271-455, 37-224	424	100% unmaskedGI	1	2	3E-46
<input type="checkbox"/>	AAC01506.1 drill through Top50BlastHits Red Seq View	AAC01506.1	type XII collagen	Homo sapiens	PRI	27.6%, 30% unmaskedSW	2-195, 2-195	133-318, 133-318	422	100% unmaskedGI	1	2	3E-50
<input type="checkbox"/>	AAC31952.1 drill through Top50BlastHits Red Seq View	AAC31952.1	Integrin subunit alpha 10 precursor	Homo sapiens	PRI	46.5%, 46% unmaskedSW	2-199, 2-199	162-359, 162-359	419	100% unmaskedGI	1	2	5E-63
<input type="checkbox"/>	CAA07569.1 drill through Top50BlastHits	CAA07569.1	matrilin-4	Homo sapiens	PRI	26%, 28% unmaskedSW	1-199, 2-200	335-529, 28-223	411	100% unmaskedGI	1	2	1E-47

THIS PAGE BLANK (USPTO)

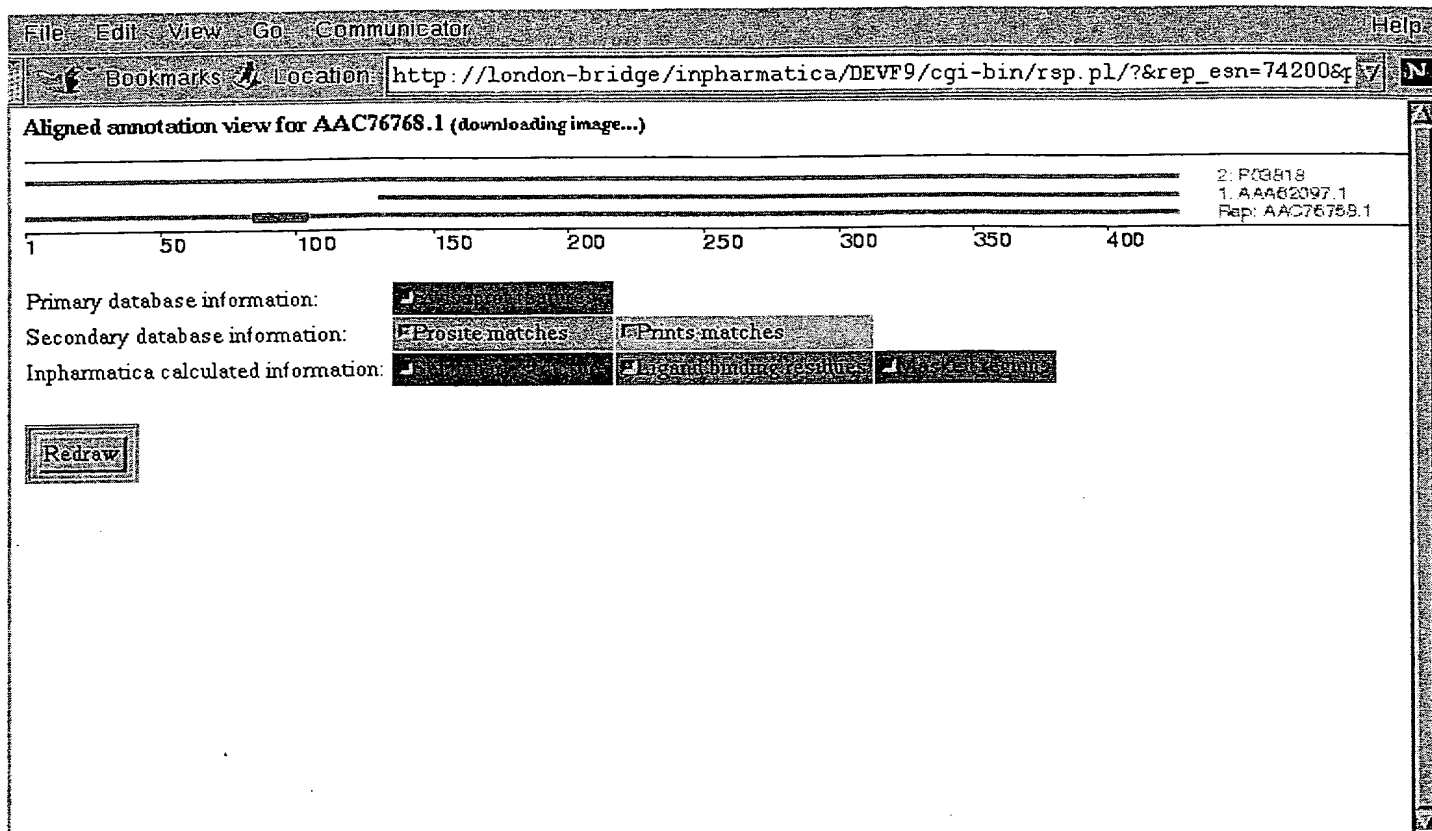
GT
confidence

□	Top50BlastHits Red. Seq. View	<u>BAA91707.1</u>	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	80-195	36-141	142	100% unmaskedGT	3	5	2E-8
□	AAA36795.1 drill through Top50BlastHits Red. Seq. View	<u>AAA36795.1</u>	undulin 2	Homo sapiens	PRI	34.7%, 21% unmaskedSW	2-46, 139-196	3-51, 37-93	121	100% unmaskedGT	2	2	8E-6
□	AAD40367.1 drill through Top50BlastHits Red. Seq. View	<u>AAD40367.1</u>	calcium-activated chloride channel-2	Homo sapiens	PRI	15%, 15% unmaskedSW	8-117, 8-117	312-421, 312-421	117	100% unmaskedGT	5	5	1E-4
□	CAA67559.1 drill through Top50BlastHits Red. Seq. View	<u>CAA67559.1</u>	collagen VI-alpha-1 chain	Homo sapiens	PRI	21.1%, 21% unmaskedSW	108-197, 108-197	7-92, 7-92	114	100% unmaskedGT	3	3	5E-7
→	AAC76768.1 drill through Top50BlastHits Red. Seq. View	<u>AAC76768.1</u>	orf, hypothetical protein	Escherichia coli	BCT	13.2%, 11% unmaskedSW	9-134, 7-186	267-384, 265-424	107	100% unmaskedGT	-4	-5	3E-24
□	CAB43000.1 drill through Top50BlastHits Red. Seq. View	<u>CAB43000.1</u>	tellurite resistance	Escherichia coli	BCT	12.9%, 14% unmaskedSW	8-192, 2-181	214-398, 208-380	102	100% unmaskedGT	-6	-7	6E-22
□	AA60114.1 drill through Top50BlastHits Red. Seq. View	<u>AA60114.1</u>	platelet membrane glycoprotein IIb	Homo sapiens	PRI	14%, 13% unmaskedSW	157-199, 157-199	145-187, 145-187	111	98.52% unmaskedGT reverse Hit	-7	-8	8E-5

THIS PAGE BLANK (USPTO)

[illegible]

THIS PAGE BLANK (USPTO)



Database information for AAC76768.1 Gene Name: yieM

Download sequence in FASTA format

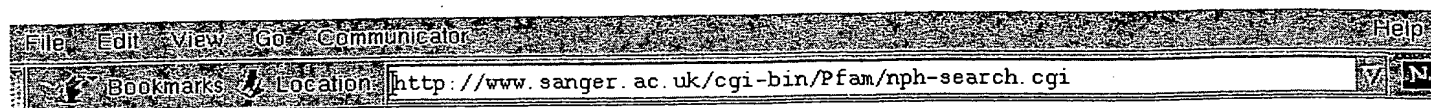
1. Name of the company	2. Name of the person	3. Name of the person	4. Name of the person
5. Name of the person	6. Name of the person	7. Name of the person	8. Name of the person
9. Name of the person	10. Name of the person	11. Name of the person	12. Name of the person

Journal references:

Journal References			
Title	Authors	Publication	Medline code
The complete genome sequence of <i>Escherichia coli</i> K-12	Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rude, C.K., Mayhew, G.E., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y.	Science 277 (5321): 1453-1474 (1997)	97426617
Direct Submission	Blattner, F.R.	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecobi@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	N/A
Direct Submission	Blattner, F.R.	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecobi@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	N/A
Direct Submission	Plunkett, G. III	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA	N/A

THIS PAGE BLANK (USPTO)

Figure 12



Pfam

Protein families database of alignments and HMMs

[Home](#) |
 [Keyword search](#) |
 [Protein search](#) |
 [DNA search](#) |
 [Browse Pfam](#) |
 [Taxonomy search](#) |
 [Help](#)

Results for gi|2367274|gb|AAC76768.1|

There were no matches to Pfam-A (including borderline matches) for gi|2367274|gb|AAC76768.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_15204	204	408	2.4e-108	Align

[427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families:

[Hypertext-linked to swisspfam](#)

Query gi|2367274|gb|AAC76768.1|/204-408 matching Pfam-B_15204

```

YIEM_ECOLI 204 DILRLPPELATLGITELEYEFYRRLVEKQLLTYYRLHGESWREKVIERPV 253
gi|2367274|gb|AAC76768.1| 204 DILRLPPELATLGITELEYEFYRRLVEKQLLTYYRLHGESWREKVIERPV 253
YIEM_ECOLI 254 VHKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMRIALAENRRCY 303
gi|2367274|gb|AAC76768.1| 254 VHKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMRIALAENRRCY 303
YIEM_ECOLI 304 IMLFSTEIVRYELSGPQGIEQAIRFLSQQFRGGTDLASCPRIMERLQSR 353
gi|2367274|gb|AAC76768.1| 304 IMLFSTEIVRYELSGPQGIEQAIRFLSQQFRGGTDLASCPRIMERLQSR 353
YIEM_ECOLI 354 EWFADADAVVISDFIAQRLPDDVT SKVKELQRVHQHMFHAVAMSANGKPGI 403
gi|2367274|gb|AAC76768.1| 354 EWFADADAVVISDFIAQRLPDDVT SKVKELQRVHQHMFHAVAMSANGKPGI 403
YIEM_ECOLI 404 MRIFD 408
gi|2367274|gb|AAC76768.1| 404 MRIFD 408

```

[Align to family](#)

If you think there is anything wrong with this script, please contact [Pfam](#)



THIS PAGE BLANK (USPTO)

Figure 13


File Edit View Go Communicator		Help
Bookmarks Location		http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=Protein
LOCUS	ARC76768 427 aa BCT 01-DEC-2000	
DEFINITION	orf, hypothetical protein [Escherichia coli K12].	
ACCESSION	ARC76768	
PID	g2367274	
VERSION	ARC76768.1 GI:2367274	
DBSOURCE	locus AE000451 accession AE000451.1	
KEYWORDS		
SOURCE	Escherichia coli K12.	
ORGANISM	Escherichia coli K12 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.	
REFERENCE	1 (residues 1 to 427)	
AUTHORS	Blattner, F.R., Plunkett, C. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, M.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.	
TITLE	The complete genome sequence of Escherichia coli K-12	
JOURNAL	Science 277 (5331), 1453-1474 (1997)	
MEDLINE	97426617	
PUBMED	9278503	
REFERENCE	2 (residues 1 to 427)	
AUTHORS	Blattner, F.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	
REFERENCE	3 (residues 1 to 427)	
AUTHORS	Blattner, F.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	
REFERENCE	4 (residues 1 to 427)	
AUTHORS	Plunkett, C. III.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA	
COMMENT	This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and HCHGR). The entire sequence was independently determined from E. coli K12 strain M61655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names. Method: conceptual translation.	
FEATURES	Location/Qualifiers	
source	1..427 /organism="Escherichia coli K12" /strain="K12" /sub_strain="M61655" /db_xref="taxon:83333"	
Protein	1..427 /function="orf; unknown" /product="orf, hypothetical protein"	
CDS	1..427 /gene="pieM" /coded_by="complement(2367272:5249..6532)" /transl_table=11 /note="f427; sequence change joins ORFs yieD and yieM from earlier version"	
ORIGIN	1 mrsrlkdarv ppetteevmc yqqsqllstp qfivqlpqil dlhlrlnspw aeqarqlvda 61 nstitsalht lflqwrsls1 ivgattlmgq lleereqll sevqermrls gqleplladn 121 ntaagrlwdm sagqlkrqdy qlivkygefl neqpelkrla eqlgrsreak siprmdagme 181 tfrtmvrepv tpeqvvgllq qsddilrlrp pelatlgite leyefyrllv ekqlltyrlh	

THIS PAGE BLANK (USPTO)

● Figure 14A

File Edit Search Type Help

Accession Code Query



Tools

←

?

?

X

Sequences

?

?

?

? Acc
y123

Ligands

? ADP
NAD
FAD

? Phe
Ala
Glu

Select a database and enter a code:

Database	
GenBank	
Accession	<input type="radio"/>
Accession Version	<input type="radio"/>
GI-NUC	<input type="radio"/>
protein_id	<input checked="" type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

AAC76768.1

Submit Reset

THIS PAGE BLANK (USPTO)

Aligned Sequence Display

Query details:

AAC76768.1:orf, hypothetical protein



Tools



Sequences



Total selected: 0

View alignment

inpharmatica Genome Threader results

Page 1 of 3

Total hits: 88

Selected: 0

Select all

Deselect all

Cluster	Detail	Code	Title	Organism	%ID	Query rgn	Target rgn	Align score	Method	Confidence
0	1	1CK4B	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN	<i>R. norvegicus</i>	12.6	267 - 398	6 - 144	158	Local	Certain (100%)
1	1	1CK4B	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN	<i>R. norvegicus</i>	19.8	267 - 398	5 - 130	157	Local	Certain (100%)
1	1	1JLM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BO...	<i>H. sapiens</i>	19.8	267 - 384	5 - 130	156	Local	Certain (100%)
1	1	1CK4A	CRYSTAL STRUCTURE OF RAT A1B1 INTEGRIN	<i>R. norvegicus</i>	12.6	267 - 398	5 - 144	149	Local	Certain (100%)
1	1	1BHO1	MAC-1 I-DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	19.8	267 - 384	5 - 130	146	Local	Certain (100%)
1	1	1BHO2	MAC-1 I-DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	19.8	267 - 384	5 - 130	146	Local	Certain (100%)
1	1	1BHO1	MAC-1 I-DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	19.8	267 - 384	5 - 130	146	Local	Certain (100%)
1	1	1IDN2	MAC-1 I-DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	19.8	267 - 384	5 - 130	146	Local	Certain (100%)
1	1	1IDN1	MAC-1 I-DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	19.8	267 - 384	5 - 130	146	Local	Certain (100%)
1	1	1BHO2	MAC-1 I-DOMAIN MAGNESIUM COMPLEX	<i>H. sapiens</i>	19.8	267 - 384	5 - 130	146	Local	Certain (100%)
1	1	1LFA-B	CD11A I-DOMAIN WITH BOUND MN++	<i>H. sapiens</i>	15.9	264 - 381	1 - 124	143	Local	Certain (100%)
1	1	1ZON	CD11A I-DOMAIN WITH BOUND MN++	<i>H. sapiens</i>	15.9	264 - 381	1 - 124	143	Local	Certain (100%)
1	1	1LFA-A	CD11A I-DOMAIN WITH BOUND MN++	<i>H. sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain (100%)
1	1	1ZOOA	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>H. sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain (100%)
1	1	1ZOPB	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>H. sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain (100%)
1	1	1ZOPA	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>H. sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain (100%)
1	1	1ZOOB	CD11A I-DOMAIN WITH BOUND MAGNESI...	<i>H. sapiens</i>	14.6	264 - 354	1 - 96	141	Local	Certain (100%)
1	1	1DGOA	NMR SOLUTION STRUCTURE OF THE INSE...	<i>H. sapiens</i>	14.3	267 - 354	10 - 100	139	Local	Certain (100%)
1	1	1ATZA	HUMAN VON WILLEBRAND FACTOR A3 DO...	<i>H. sapiens</i>	11.3	267 - 407	6 - 152	135	Local	Certain (100%)
1	1	1OAKA	CRYSTAL STRUCTURE OF THE VONWILLE...	<i>H. sapiens</i>	11.3	267 - 398	10 - 150	131	Local	Certain (100%)
1	1	1AOXA	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>H. sapiens</i>	13.2	267 - 384	9 - 134	107	Local	Certain (100%)
1	1	1QC5B	I-DOMAIN FROM INTEGRIN ALPHA1-BETA1	<i>N/A</i>	11.9	267 - 398	5 - 143	106	Local	Certain (100%)
1	1	1QC5A	I-DOMAIN FROM INTEGRIN ALPHA1-BETA1	<i>N/A</i>	11.9	267 - 398	7 - 145	102	Local	Certain (100%)
1	1	1AOXB	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1	<i>H. sapiens</i>	13.8	267 - 398	11 - 149	98	Local	Certain (100%)
1	1	1AO3A	A3 DOMAIN OF VON WILLEBRAND FACTOR	<i>H. sapiens</i>	10.7	267 - 398	8 - 145	98	Local	Certain (100%)
1	1	1AO3B	A3 DOMAIN OF VON WILLEBRAND FACTOR	<i>H. sapiens</i>	10.7	267 - 398	8 - 145	98	Local	Certain (100%)
1	1	1AUQ	A1 DOMAIN OF VON WILLEBRAND FACTOR	<i>H. sapiens</i>	11.8	250 - 398	4 - 159	97	Local	Certain (100%)
1	1	1ATZB	HUMAN VON WILLEBRAND FACTOR A3 DO...	<i>H. sapiens</i>	10.7	267 - 398	9 - 146	96	Local	Certain (100%)
1	1	1A6JB	NITROGEN REGULATORY BACTERIAL PRO...	<i>E. coli</i>	19.6	306 - 349	13 - 57	60	Local	Low (42%)
1	1	1BKDS	COMPLEX OF HUMAN H-RAS WITH HUMA...	<i>H. sapiens</i>	14.4	141 - 237	100 - 202	69	Local	Low (34%)

THIS PAGE BLANK (USPTO)

Aligned Sequence Display

Query details:
AAC76768.1:orf, hypothetical protein



Tools									
<div> <div>File Edit Search Type Help</div> <div> <div>Page 1 of 17</div> <div> <div>Total hits: 508</div> <div>Selected: 0</div> </div> </div> <div> <div>Reverse Maximized PSI-BLAST results</div> <div>View alignment</div> </div> </div>									
Cluster	Detail	Code	Title	Organism	%ID	Query idn	Target idn	Bit	E-value
o	o	BAA80016.1	452aa long hypothetical protein	<i>A. permix</i>	28.0	156 - 407	202 - 431	2	2.00E-60
o	o	Q60384	HYPOTHETICAL PROTEIN M30077	<i>Mjannaschii</i>	22.0	146 - 408	122 - 361	3	2.00E-71
o	o	AAB8810.1	conserved hypothetical protein	<i>Mjannaschii</i>	23.0	147 - 410	170 - 420	3	2.00E-65
o	o	AAG79610.1	putative retroelement polypolyprotein	<i>Athaliana</i>	19.0	168 - 404	144 - 403	3	2.00E-50
o	o	Q55874	HYPOTHETICAL 45.8 KDA PROTEIN SL10103	<i>S.sp.</i>	21.0	241 - 407	17 - 188	3	4.00E-39
o	o	AAB84857.1	magnesium chelatase subunit Chl	<i>Mithermotrophicum</i>	25.0	151 - 370	324 - 535	3	5.00E-22
o	o	CAA30478.1	Not given	<i>Mmusculus</i>	20.0	247 - 416	132 - 309	4	2.00E-44
o	o	AAAS5544.1	Not given	<i>Hsapilens</i>	16.0	223 - 412	107 - 305	4	3.00E-42
o	o	AAB24821.1	leukocyte integrin alpha chain	<i>Hsapilens</i>	16.0	223 - 412	107 - 305	4	3.00E-42
o	o	1JLMR	I-DOMAIN FROM INTEGRIN CR3, MN2+ BOUND	<i>Hsapilens</i>	16.0	263 - 412	11 - 158	4	2.00E-41
o	o	1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BOUND	<i>Hsapilens</i>	16.0	263 - 412	11 - 158	4	2.00E-41
o	o	1BHO1	MAG-1 DOMAIN MAGNESIUM COMPLEX	<i>Hsapilens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	o	1IDN2	MAC-1 DOMAIN METAL FREE	<i>Hsapilens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	o	1IDN1	MAG-1 DOMAIN METAL FREE	<i>Hsapilens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	o	1BHO2	MAC-1 DOMAIN MAGNESIUM COMPLEX	<i>Hsapilens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	o	1BHO2	MAG-1 DOMAIN CADMIUM COMPLEX	<i>Hsapilens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	o	1BHO1	MAC-1 DOMAIN CADMIUM COMPLEX	<i>Hsapilens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	o	1BHO1	MAGNESIUM CHELATASE SUBUNIT CHLIPRECURSOR	<i>Hsapilens</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	o	Q22437	Not given	<i>unidentified</i>	18.0	162 - 417	473 - 705	4	2.00E-40
o	o	CAB69268.1	Not given	<i>Ssacifa</i>	17.0	254 - 417	6 - 177	4	8.00E-40
o	o	AAB16669.1	CD11b	<i>Hvulgare</i>	17.0	162 - 410	90 - 311	4	2.00E-38
o	o	CAB58179.1	Mg-protoporphyrin IX	<i>Athaliana</i>	17.0	162 - 423	441 - 685	4	5.00E-38
o	o	AAD52081.1	magnesium chelatase subunit CHL D precursor	<i>S.sp.</i>	18.0	199 - 367	415 - 588	4	6.00E-38
o	o	BAA16787.1	Mg chelatase subunit	<i>Mithermotrophicum</i>	13.0	154 - 393	50 - 303	4	7.00E-37
o	o	AAB66281.1	conserved protein	<i>Rcapsulatus</i>	21.0	168 - 367	277 - 484	4	1.00E-36
o	o	CAA77537.1	561 aa (60 kD) Mg chelatase subunit	<i>Athaliana</i>	17.0	162 - 423	491 - 747	4	2.00E-36
o	o	AAB22885.1	T27G720	<i>Rnorvegicus</i>	14.0	239 - 412	126 - 307	4	2.00E-35
o	o	AAF21241.1	alpha D integrin	<i>Mmusculus</i>	13.0	231 - 412	118 - 307	4	2.00E-35
o	o	AAF23482.1	leukocyte adhesion glycoprotein gp150/95 alpha integrin su	<i>Hsapilens</i>	13.0	238 - 410	124 - 303	4	3.00E-35
o	o	AAB38547.1	leukointegrin alpha d chain						

THIS PAGE BLANK (USPTO)

Figure 15

AlEye output (December 13, 2000 3:07 PM)

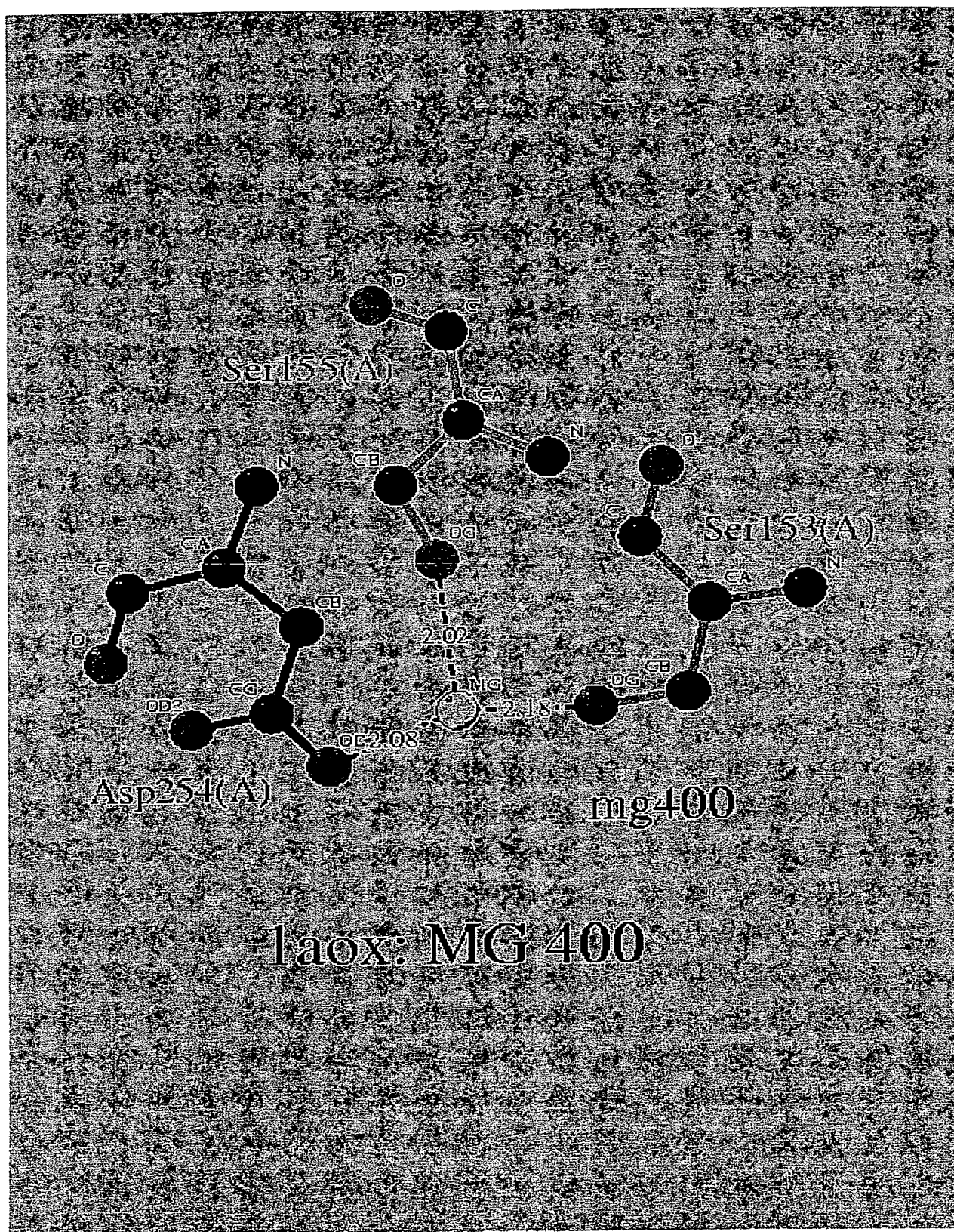
```

AAC78768.1      10      20      30      40      50      60      70
1AOX:A      MRSRLKDA RVPPPEL TEEVMCYQQS QLLSTPQ FIVQLPQ ILDLHRLNSPWA EQARQLVDANSTITSA LHTLFLQRPWR
-----
AAC78768.1      80      90      100      110      120      130      140      150
1AOX:A      LSLIVQATT LNQQLLEEREQ LLSLVQERMTLSGQLEPI LADNNTAAGRLWDM SAGQLKRGDYQLIVKYGEFLNEQP
-----
AAC78768.1      160      170      180      190      200      210      220      230
1AOX:A      ELKRLAEQLGRSREAKSIPRND AQMETFTRTMVREPA TVPEQVDGLQQSDDILRLLPPELATLGIT ELEYEFYRRLVE
-----
AAC78768.1      240      250      260      270      280      290      300
1AOX:A      KQLLT YRLHGESWREKVIERPPV VHKDYDEQPRGPFIVCVDTSGSMGGFNEQCAKAFCLALMPI--ALAE NRRCYIML
-----
AAC78768.1      310      320      330      340      350      360      370      380
1AOX:A      FSTEI--VRYELSGPQG---IEQA I RFLSQQFRGGTDLASC FRAIMERLQS-----REWFDADA VVISDFIAQRLPDD
YANNPRVVFNLT YKTKEEMIVAT SQTSQYGGDLINTFGA IQYARKYAYSAASGGRRSATKVMVVVYTDGESH-DGSM
390      400      410      420      430      440      450      460
AAC78768.1      VTSKVVKELQ RVHQHRRFHAVAMSAHGKPGIMRIFDH IWRFDTGMPRSRLLRWRP
1AOX:A      LKAVIDQCNHON-ILRFGIAVLGY-----LN RNALDTKNLIK EIKAIASIP TERYFFNFVSD EAAALLEKAG
470
AAC78768.1
1AOX:A      TLGEQIFSIEGGT

```

THIS PAGE BLANK (USPTO)

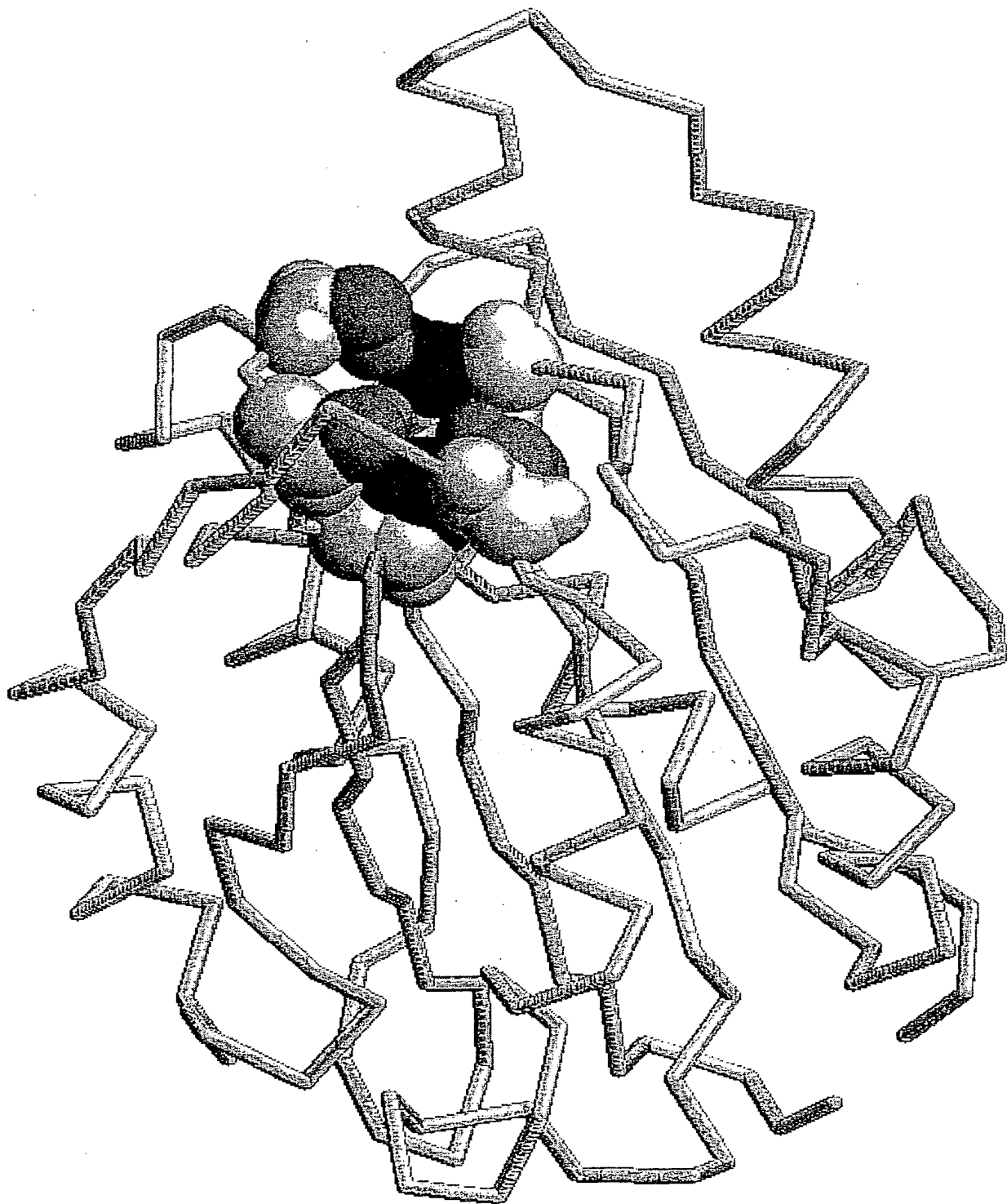
Figure 16A



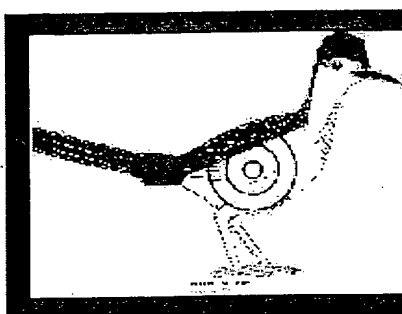
THIS PAGE BLANK (USPTO)

● gne.16B

C



THIS PAGE BLANK (USPTO)



Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:

☐ Homo sapiens ☐ Rattus norvegicus (Rat) ☐ Mus musculus (Mouse) ☐ Danio rerio (Zebrafish)

THIS PAGE BLANK (USPTO)

File Edit View Go Communicator

Bookmarks Location http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl#line2

2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!

Add2hit	BPD links	WAVY link	Title	Organism	Div.	%ID (GI, PSI)	Query rgn. (GI, PSI)	Target rgn. (GI, PSI)	Aln. score (GI)	Conf. (GI)	1st Iter. (PSI)	Best Iter. (PSI)	Best E-value (PSI)
	AAA59544.1 drill through Top50BlastHits Red. Seq. View	AAA59544.1	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGI	1	1	1E-104
	AA824821.1 drill through Top50BlastHits Red. Seq. View	AA824821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGI	1	1	1E-104
	P99715 drill through Top50BlastHits Red. Seq. View	Q99715	COLLAGEN ALPHA (XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.9%, 28% unmaskedSW	2-186, 2-179	439-617, 2322-2494	456	100% unmaskedGI	1	2	1E-49
	AA838702.1 drill through Top50BlastHits Red. Seq. View	AA838702.1	cartilage matrix protein	Homo sapiens	PRI	28.9%, 25% unmaskedSW	2-186, 2-186	274-452, 40-221	446	100% unmaskedGI	1	2	1E-48
	AAC01508.1 drill through Top50BlastHits Red. Seq. View	AAC01508.1	type XII collagen	Homo sapiens	PRI	28.4%, 28% unmaskedSW	2-186, 2-186	137-318, 137-318	445	100% unmaskedGI	1	2	1E-56
	CAA72402.1 drill through Top50BlastHits Red. Seq. View	CAA72402.1	collagen type XIV	Homo sapiens	PRI	28.7%, 30% unmaskedSW	2-186, 2-186	6-185, 6-185	442	100% unmaskedGI	1	2	3E-57
	AA838547.1 drill through Top50BlastHits Red. Seq. View	AA838547.1	leukointegrin alpha d chain	Homo sapiens	PRI	61%, 60% unmaskedSW	1-187, 1-187	148-334, 148-334	439	100% unmaskedGI	1	2	3E-66
	CA871222.1 drill through Top50BlastHits Red. Seq. View	CA871222.1	dJ238015.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	27.1%, 22% unmaskedSW	1-186, 2-186	293-472, 1430-1620	439	100% unmaskedGI	1	3	6E-25
	CA407589.1 drill through Top50BlastHits Red. Seq. View	CAA07569.1	matrilin-4	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-187	344-525, 34-215	418	100% unmaskedGI	2	2	2E-45
	CA848380.1 drill through Top50BlastHits Red. Seq. View	CA848380.1	dJ453C12.3 (matrilin-4)	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-186	385-566, 34-214	418	100% unmaskedGI	2	2	5E-45
	CA870853.1 drill through Top50BlastHits Red. Seq. View	CA870853.1	hypothetical protein	Homo sapiens	PRI	25.9%, 26% unmaskedSW	3-186, 3-186	440-620, 440-620	408	100% unmaskedGI	1	2	7E-54
	AAT43107.1 drill through Top50BlastHits Red. Seq. View	AAT43107.1	HUMINAE	Homo sapiens	PRI	37.5%, 38% unmaskedSW	6-187, 6-187	153-336, 153-336	402	100% unmaskedGI	1	2	4E-55
	P38520 drill through Top50BlastHits Red. Seq. View	P38520	INTEGRIN ALPHA-E PRECURSOR (MUCOSAL LYMPHOCYTE-1 ANTIGEN) (HML-1 ANTIGEN) (CD103 ANTIGEN) (INTEGRIN ALPHA-IEL).	Homo sapiens (Human).	PRI	37.5%, 38% unmaskedSW	6-187, 6-187	205-388, 205-388	402	100% unmaskedGI	1	2	4E-55
	P56199 drill through Top50BlastHits Red. Seq. View	P56199	INTEGRIN ALPHA-1 (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD49A).	Homo sapiens (Human).	PRI	30.1%, 31% unmaskedSW	3-187, 3-187	144-333, 144-333	395	100% unmaskedGI	1	2	8E-59

THIS PAGE BLANK (USPTO)

Figure 18B

http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl?Mine2

1) 160 hits identified by Genome Threader only:

Add3Hit	Redundant sequence display	BFD link	WWW link	Title	Organism	Div.	%ID	Query len.	Target len.	Sim. score	Conf.
<input type="checkbox"/>	Red. Seq. View	CAA82910.1 drill through Top50BlastHits	CAA82910.1	basic transcription factor 2, 44 kD subunit	Homo sapiens	PRI	12.4% unmaskedSW	4-169	61-224	100	100% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	BAA30761.1 drill through Top50BlastHits	BAA30761.1	Not given	Homo sapiens	PRI	12.9% unmaskedSW	4-183	11636-2036	97	100% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAA36154.1 drill through Top50BlastHits	AAA36154.1	Not given	Homo sapiens	PRI	11.5% unmaskedSW	4-186	10-235	87	100% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	CAA15889.1 drill through Top50BlastHits	CAA15889.1	αJ745E6 I (Breast Cancer Suppressor Candidate I (BCSC-1) LIKE)	Homo sapiens	PRI	15.8% unmaskedSW	68-181	6-107	90	99.93% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAF71133.1 drill through Top50BlastHits	AAF71133.1	PR02768	Homo sapiens	PRI	14.7% unmaskedSW	108-179	1-71	100	98.52% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAB60942.1 drill through Top50BlastHits	AAB60942.1	breast cancer suppressor candidate 1	Homo sapiens	PRI	19.9% unmaskedSW	62-167	2-95	83	97.43% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAF03046.1 drill through Top50BlastHits	AAF03046.1	candidate tumor suppressor protein DICE1	Homo sapiens	PRI	17.7% unmaskedSW	4-112	4-131	81	97.43% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	CAB62520.1 drill through Top50BlastHits	CAB62520.1	hypothetical protein	Homo sapiens	PRI	26.8% unmaskedSW	73-143	73-142	73	93.85% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAA61177.1 drill through Top50BlastHits	AAA61177.1	Not given	Homo sapiens	PRI	9.4% unmaskedSW	4-161	38-228	68	93.33% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	BAA31611.1 drill through Top50BlastHits	BAA31611.1	KIA0636 protein	Homo sapiens	PRI	20.2% unmaskedSW	73-185	390-503	66	92.74% unmaskedGI reverse HI
<input type="checkbox"/>	Red. Seq. View	CAB52192.1 drill through Top50BlastHits	CAB52192.1	G7c protein	Homo sapiens	PRI	14.1% unmaskedSW	6-88	20-105	76	89.66% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	BAA75899.1 drill through Top50BlastHits	BAA75899.1	N-copine	Homo sapiens	PRI	26.8% unmaskedSW	73-143	405-474	73	89.66% unmaskedGI
<input checked="" type="checkbox"/>	Red. Seq. View	AAD05645.1 drill through Top50BlastHits	AAD05645.1	putative	Helicobacter pylori J99	BCT	13.4% unmaskedSW	1-157	342-514	68	89.66% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	P26012 drill through Top50BlastHits	P26012	INTEGRIN BETA-8 PRECURSOR.	Homo sapiens (Human).	PRI	13.1% unmaskedSW	2-115	147-289	78	97.73% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAD21820.1 drill through Top50BlastHits	AAD21820.1	NC37	Homo sapiens	PRI	14.1% unmaskedSW	6-88	318-403	76	85.59% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	CAA65775.1 drill through Top50BlastHits	CAA65775.1	Sec23 protein	Homo sapiens	PRI	19.7% unmaskedSW	7-141	134-275	74	84.45% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAC15920.1 drill through Top50BlastHits	AAC15920.1	copine I	Homo sapiens	PRI	20% unmaskedSW	73-185	384-497	67	80.73% unmaskedGI reverse HI
<input type="checkbox"/>	Red. Seq. View	AAD06777.1 drill through Top50BlastHits	AAD06777.1	TRYPTOPHAN SYNTHASE ALPHA CHAIN	Helicobacter pylori J99	BCT	11.8% unmaskedSW	16-185	30-198	44	80.79% unmaskedGI
<input type="checkbox"/>	Red. Seq. View	AAD06358.1 drill through Top50BlastHits	AAD06358.1	putative SUGAR NUCLEOTIDE BIOSYNTHESIS PROTEIN	Helicobacter pylori J99	BCT	16.7% unmaskedSW	107-186	13-84	63	74.2% unmaskedGI

THIS PAGE BLANK (USPTO)

626 out of these 649 PSI-BLAST matches were identified using 'positive iterations':

[illegible]

23 out of these 649 PSI-BLAST matches were identified using 'negative iterations':

[illegible]

A2) Genome Threader Matches:

2925 matches found by Genome Threading:

[illegible]

THIS PAGE BLANK (USPTO)

Figure 19

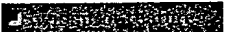
File Edit View Go Communicator Help



Bookmarks Location: http://london-bridge/inpharmatica/DEVF9/cgi-bin/rsp.pl/?&rep_esn=85522&




Aligned annotation view for AAD05645.1 (downloading image...)


Rep: AAD05645.1

1 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800

Primary database information: 


Secondary database information:  


Inpharmatica calculated information:   





Database information for AAD05645.1 Gene Name: [Dp0061](#)


Download sequence in FASTA format

















































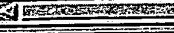















Journal references:

Title	Authors	Publication	Medline code
Genomic sequence comparison of two unrelated isolates of the human gastric pathogen <i>Helicobacter pylori</i>	Alm R.A., Ling G., S.T., Morad J., King B.L., Brown E.D., Dong P.C., Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carniel G., Tummino P.J., Garuso A., Una-Nikelsen M., Mills D.M., Ives C., Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F. and Trust T.J.	Nature 397 (6715): 176-180 (1999)	99120557
Direct Submission:	King B.L., Alm R.A. and Trust T.J.	Submitted (12 JAN 1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA	N/A

External database: GI Nucleotide identifier from the NCBI [4154569](#)

External database: NCBI database of taxonomy [85963](#)

100%

THIS PAGE BLANK (USPTO)

que 20

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: <http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi>



Pfam

Protein families database of alignments and HMMs

[Home](#) |
 [Keyword search](#) |
 [Protein search](#) |
 [DNA search](#) |
 [Browse Pfam](#) |
 [Taxonomy search](#) |
 [Help](#)



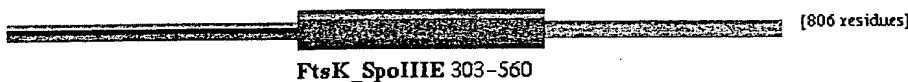
Results for gi|4154569|gb|AAD05645.1|

Trusted matches - domains scoring higher than the gathering threshold

Domain	Start	End	Bits	Evalue	Alignment
FtsK_SpoIIIE	303	560	208.60	9.4e-59	Align

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_38662	1	302	3e-158	Align
Pfam-B_36020	561	806	3.2e-129	Align



Alignments of Pfam-A domains to HMMs

Format for fetching alignments to seed [Hypertext-linked to swisspfam](#)

Alignment of FtsK_SpoIIIE vs gi|4154569|gb|AAD05645.1|/303-560

```

*->elkelldskafdrsrslrtialgkdivsgnpvvadlvkmpvdaqprH
gi|4154569 303  elk+l+++k+f++ +s+ ++ +++ ++++++v++ ++ ++ M
ELKDLQRDKKEFWTKSSSQHEVSVPVGVWDINHKKEVCFKIGNEQN---M 345

LIaGatGSGKSvflntliIsIaaxhsPeeVrlyliDpkGg.eLapledl
gi|4154569 346 LI+ +GSGKS+fl+ li++la++++p+eV+l+l+D+K+g e++++
TLICDHSGSGKSNFLHVLIQNLAFYYDPDEVQLFLLDYKEGVFNAYVAD 395

phllsv.lvavatdpekAlralreLveEMerRyerallrqlgvrsieey.
gi|4154569 396 p l +++lv v++++ + +l++L++EM++R + ++q++v+++++p+
PALEHARLVSVASSISYGITFLKWLCEDEMQRA--DRFKQFNVKDLSDYr 443

yneciaevilggfgdvdlviDnydeLaeenLlervtsLknqGlsygvhvm
gi|4154569 444 +++e++++ i+++++++l++Dn++++a+e+ +++t+Lk+++++++ v+
KHEKMPRLIVVIDEFQVLFSDNKSTKAVEG--HLNTLLKKGRSYGVHLVL 491

atagrwl.eLpyivvVDeaeLmlaagkdseLRsrvedllaRlaqmk
gi|4154569 492 at+++++++p+ +++ +r++l ++++ s ++ +d+ +++ +
RTQTRGTdNPSFKAQIANRIALPMDREDSS---SVLGDDARCEIQKPE 538

eeltfaRlnlDdaaciMlllatpeldaqpDvavvqRpsv+-*
gi|4154569 539 + +n ++G ++t+++ +
-----GTFN---NNGGNRKYHTKMS-----VPKAP 560

```

[Align to seed](#)

THIS PAGE BLANK (USPTO)

File Edit View Go Communicator Help

Bookmarks Location <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retr> N

NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM

Search Protein for I Go Clear

Links Index History Clipboard

Display Default View as HTML Save Add to Clipboard Hide Brief and LinkBox

1: GI "4154569" [GenPept] putative [Helicobacter pylori... BLink, PubMed, Related Sequences, Nucleotide, Genome, Taxonomy

LOCUS AAD05645 806 aa BCT 20-JAN-1999

DEFINITION putative [Helicobacter pylori J99].

ACCESSION AAD05645

PID g4154569

VERSION AAD05645.1 GI:4154569

DBSOURCE locus AE001445 accession AE001445.1

KEYWORDS

SOURCE Helicobacter pylori J99.

ORGANISM Helicobacter pylori J99
Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group; Helicobacter.

REFERENCE 1 (residues 1 to 806)
AUTHORS Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelsen, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F. and Trust, T.J.
TITLE Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori
JOURNAL Nature 397 (6715), 176-180 (1999)
MEDLINE 99120557
REMARK Erratum: [[published erratum appears in Nature 1999 Feb 25; 397(6721):719]]

REFERENCE 2 (residues 1 to 806)
AUTHORS King, B.L., Alm, R.A. and Trust, T.J.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney Street, Cambridge, MA 02139, USA

COMMENT Method: conceptual translation supplied by author.

FEATURES

source 1..806
/organism="Helicobacter pylori J99"
/strain="J99"
/db_xref="taxon:85963"

Protein 1..806
/product="putative"

CDS 1..806
/gene="jhp0061"
/coded_by="4154559:5018..7438"
/transl_table=11
/note="similar to H. pylori 26695 gene HP0066"

ORIGIN

```

1 mieisewlqk lddaldkvva kkepesflkp iispiedyqk svrqiaqft dapkfneega
61 ypqflscgll qvrgknganm eflpkvypf ppkslyiehe kdgqflrem1 mrlssaplv
121 qlevilidal slggifnlar rlldkndfi yqqriltesk eieealkhlh eplkvnlqek
181 lagfrdfvhy nenakdslpl kalflsgvda lskdalyyle kimrfgskng vlsfvmlese
241 knnqsaedlk ryaeffkdrf sfecdkylnv eiisdqgiks qhmqdfadki kappkqkkev
301 krelkdldqr kefwtkssqh evsvpvgwdi nhkevcfkig neqhtliod hsgsgksnfl
361 hvliqnlafy ydpdevqlfl ldykegvfev ayvadpaleh arlvsvassi sygitflkw1
421 cdemqkradr fkqfnvkdls dyrkhekmpv livvidefqv lfsdnkstka veghlntllk
481 kgrsygvhlv latqtnrgtd inpsfkagia nrialpmdae dsssvlgdda aceiqkpegi
541 fnnnggnrky htkmsvpkap ddfksiltki haefnqnrla pidrkiyng tplkmpdtlk
601 anemrlhlgl kvdyeqkdli vefesneshl lvviqdlrar ialmkllfqv vksankelvf
661 cnkekrlirs fdaqkevgit pvenilsvld tamnpnsalv idnlneake1 hdkvgaeklk
721 sflekaidne qvovifahdf rqihtnyhfd klkellnnhf kqolafrore enlnaiksdl
781 pppsklnvll ielskdsvte frpfsl


```

//







THIS PAGE BLANK (USPTO)

File Edit Search Type Help





Accession Code Query





Tools

Sequences

Ligands

Select a database and enter a code:

Database	Field	Search
GenBank	Accession	<input type="radio"/>
	Accession Version	<input type="radio"/>
	GI-NID	<input type="radio"/>
	protein_id	<input checked="" type="radio"/>
	GI-PID	<input type="radio"/>
SWISS-PROT	Accession	<input type="radio"/>
	ID	<input type="radio"/>
PDB	ID	<input type="radio"/>



THIS PAGE BLANK (USPTO)

Aligned Sequence Display

Query details:

AAD05645.1: putative



Figure 22B

Tools: Sequences: Ligands:

Page 1 of 7 Total hits: 186 Selected: 0

Total selected: 0 View alignment

Inpharmatica Genome Threader results

Cluster	Details	Code	Title	Organism	% ID	Query Yr	Target Yr	Align. score	Method	Confidence
→	1JLM	I-DOMAIN FROM INTEGRIN CR3, MN2+ BO...	<i>Hsapiens</i>	13.4	342 - 514	1 - 157	68	Local	Low (89%)	
	1E0J.B	GP4D HELICASE FROM PHAGE T7 ADPNP ...	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1E0J.C	GP4D HELICASE FROM PHAGE T7 ADPNP ...	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1E0J.E	GP4D HELICASE FROM PHAGE T7 ADPNP ...	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1E0J.E	GP4D HELICASE FROM PHAGE T7 ADPNP ...	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1E0K.A	GP4D HELICASE FROM PHAGE T7	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1E0K.C	GP4D HELICASE FROM PHAGE T7	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1E0K.E	GP4D HELICASE FROM PHAGE T7	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1E0K.F	GP4D HELICASE FROM PHAGE T7	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1E0K.D	GP4D HELICASE FROM PHAGE T7	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1E0K.B	GP4D HELICASE FROM PHAGE T7	N/A	11.6	346 - 503	47 - 213	73	Local	Low (88%)	
	1A5T	CRYSTAL STRUCTURE OF THE DELTA PR...	<i>E.coli</i>	10.2	344 - 514	25 - 312	71	Local	Low (86%)	
	1BHO.1	MAC-1 I DOMAIN MAGNESIUM COMPLEX	<i>Hsapiens</i>	13.8	347 - 514	6 - 157	62	Local	Low (85%)	
	1BHO.1	MAC-1 I DOMAIN MAGNESIUM COMPLEX	<i>Hsapiens</i>	13.8	347 - 514	6 - 157	62	Local	Low (85%)	
	11DN.2	MAC-1 I DOMAIN METAL FREE	<i>Hsapiens</i>	13.8	347 - 514	6 - 157	62	Local	Low (85%)	
	1BHO.2	MAC-1 I DOMAIN MAGNESIUM COMPLEX	<i>Hsapiens</i>	13.8	347 - 514	6 - 157	62	Local	Low (85%)	
	11DN.1	MAC-1 I DOMAIN METAL FREE	<i>Hsapiens</i>	13.8	347 - 514	6 - 157	62	Local	Low (85%)	
	1BHO.2	MAC-1 I DOMAIN MAGNESIUM COMPLEX	<i>Hsapiens</i>	13.8	347 - 514	6 - 157	62	Local	Low (85%)	
	1E0J.A	GP4D HELICASE FROM PHAGE T7 ADPNP ...	N/A	12.4	346 - 500	46 - 209	69	Local	Low (83%)	
	1E0J.D	GP4D HELICASE FROM PHAGE T7 ADPNP ...	N/A	12.4	346 - 500	46 - 209	69	Local	Low (83%)	
	1IDO	I-DOMAIN FROM INTEGRIN CR3, MG2+ BO...	<i>Hsapiens</i>	14.0	342 - 535	1 - 178	59	Local	Low (80%)	
	1EPP.A	PROTEIN FARNESYL TRANSFERASE COMPL...	<i>Rhodospirillum rubrum</i>	9.7	96 - 268	187 - 261	55	Local	Low (74%)	
	1DDE.A	STRUCTURE OF THE DNAG CATALYTIC CO...	<i>E.coli</i>	15.8	569 - 746	78 - 266	63	Local	Low (70%)	
	1UAA.B	STRUCTURE OF THE REP HELICASE SINGL...	<i>E.coli</i>	13.8	106 - 257	59 - 208	75	Local	Low (62%)	
	1EQN.B	E. COLI PRIMASE CATALYTIC CORE	N/A	16.2	569 - 756	81 - 281	59	Local	Low (59%)	
	1EQN.E	E. COLI PRIMASE CATALYTIC CORE	N/A	15.8	569 - 745	81 - 270	62	Local	Low (56%)	
	1EQN.C	E. COLI PRIMASE CATALYTIC CORE	N/A	16.2	569 - 756	82 - 282	58	Local	Low (56%)	
	1VOLA.1	TFIIB (HUMAN CORE DOMAIN) TBP (A. THA...	<i>Hsapiens</i>	8.0	104 - 176	101 - 187	57	Local	Low (56%)	
	1EFJB	HOMOLOGY MODEL OF THE MOUSE RVRL...	N/A	21.7	643 - 767	51 - 181	61	Local	Low (55%)	
	1EQN.A	E. COLI PRIMASE CATALYTIC CORE	N/A	16.2	569 - 756	83 - 289	59	Local	Low (55%)	

GenBank | SWISS-PROT | PDB

THIS PAGE BLANK (USPTO)

Aligned Sequence Display

Query details:

AAD05645.1: putative



Figure 22c

Tools: Sequences: Ligands:

Page 1 of 6
Total hits: 171
Selected: 0

Total selected: 0
Reverse Maximized PSI-BLAST results
View alignment

Cluster	Detail	Code	Title	Organism	% ID	Query rgn	Target rgn	Bit	E value
o	conserved hypothetical ATP-binding protein	AAD07128.1		<i>H. pylori</i>	82.0	1 - 806	24 - 831	1	0.00E+00
o	cell division-related gene	BAA33499.1		<i>H. pylori</i>	76.1	320 - 515	155 - 352	1	1.00E-73
o	cell division protein (ftsK)	AAD08132.1		<i>H. pylori</i>	23.0	321 - 628	521 - 832	2	6.99E-96
o	septum formation protein	AAD05914.1		<i>H. pylori</i>	24.0	321 - 628	507 - 818	2	2.00E-95
o	Not given	CAA53289.1		<i>C. burnetii</i>	22.0	322 - 636	409 - 757	2	6.00E-83
o	CELL DIVISION PROTEIN FTSK HOMOLOG (ftsK)	CAA15248.1		<i>R. prowazekii</i>	20.0	322 - 644	376 - 740	2	5.00E-80
o	cell division protein FtsK, putative	AAF09880.1		<i>Dracodurans</i>	23.0	321 - 638	626 - 954	2	2.00E-77
o	cell division protein FtsK	AAF41689.1		<i>N. meningitidis</i>	19.0	321 - 646	450 - 808	2	1.00E-73
o	Not given	CAB84755.1		<i>N. meningitidis</i>	18.0	321 - 646	405 - 763	2	2.00E-69
o	Cell Division Protein FtsK	AAD19018.1		<i>C. pneumoniae</i>	20.0	321 - 581	455 - 725	2	1.00E-68
o	Not given	CAB15175.1		<i>B. subtilis</i>	30.0	322 - 566	371 - 620	2	5.00E-65
o	cell division protein FtsK-related protein	AAC23240.1		<i>H. influenzae</i>	20.0	370 - 646	211 - 525	2	4.00E-62
o	Not given	AAA61512.1		<i>C. jejuni</i>	24.0	321 - 533	604 - 814	2	5.00E-27
o	Not given	AAA22784.1		<i>B. subtilis</i>	30.0	321 - 395	439 - 511	2	2.00E-23
o	FtsK-like protein	CAA63241.1		<i>A. brasilense</i>	17.0	370 - 456	515 - 631	2	4.00E-19
o	cell division protein putative	AAC66637.1		<i>B. burgdorferi</i>	31.0	312 - 492	498 - 553	2	3.00E-14
o	putative cell division protein	CAB73144.1		<i>C. jejuni</i>	25.0	370 - 410	657 - 692	2	3.00E-07
o	Not given	CAB15175.1		<i>B. subtilis</i>	21.0	431 - 582	806 - 933	2	4.00E-04
o	hypothetical protein Rv3447c	CAB08677.1		<i>M. tuberculosis</i>	22.0	322 - 717	397 - 863	3	5.00E-65
o	Not given	CAA20269.1		<i>S. coloratus</i>	23.0	322 - 621	441 - 767	3	3.00E-58
o	hypothetical protein	CAA75203.1		<i>M. leprae</i>	23.0	322 - 560	443 - 705	3	8.99E-58
o	hypothetical protein Rv3670	CAA17962.1		<i>M. tuberculosis</i>	25.0	322 - 550	445 - 694	3	4.00E-57
o	hypothetical protein Rv3894c	CAB08080.1		N/A	23.0	324 - 601	492 - 783	3	2.00E-55
o	hypothetical protein Rv0284	CAB17359.1		<i>M. tuberculosis</i>	19.0	321 - 630	445 - 771	3	2.00E-51
o	Not given	AAA88410.1		<i>P. putida</i>	21.0	325 - 601	267 - 546	3	5.00E-44
o	Not given	BAA19323.1		<i>B. subtilis</i>	19.0	341 - 594	225 - 477	3	4.00E-38
o	putative membrane protein	CAB61675.1		<i>S. coloratus</i>	19.0	344 - 593	271 - 512	3	1.00E-37
o	Trab	AAAG1013.1		<i>S. phaeochromogenes</i>	20.0	354 - 523	276 - 449	3	2.00E-21
o	Not given	CAA56759.1		<i>S. ghanensis</i>	23.0	443 - 532	487 - 576	3	9.00E-18
o	514aa long hypothetical protein	BAA81104.1		<i>A. permix</i>	17.0	446 - 540	401 - 489	3	2.00E-11

GenBank SWISS-PROT PDB



AI Eye output (December 18, 2000 3:12 PM)

Figure 23

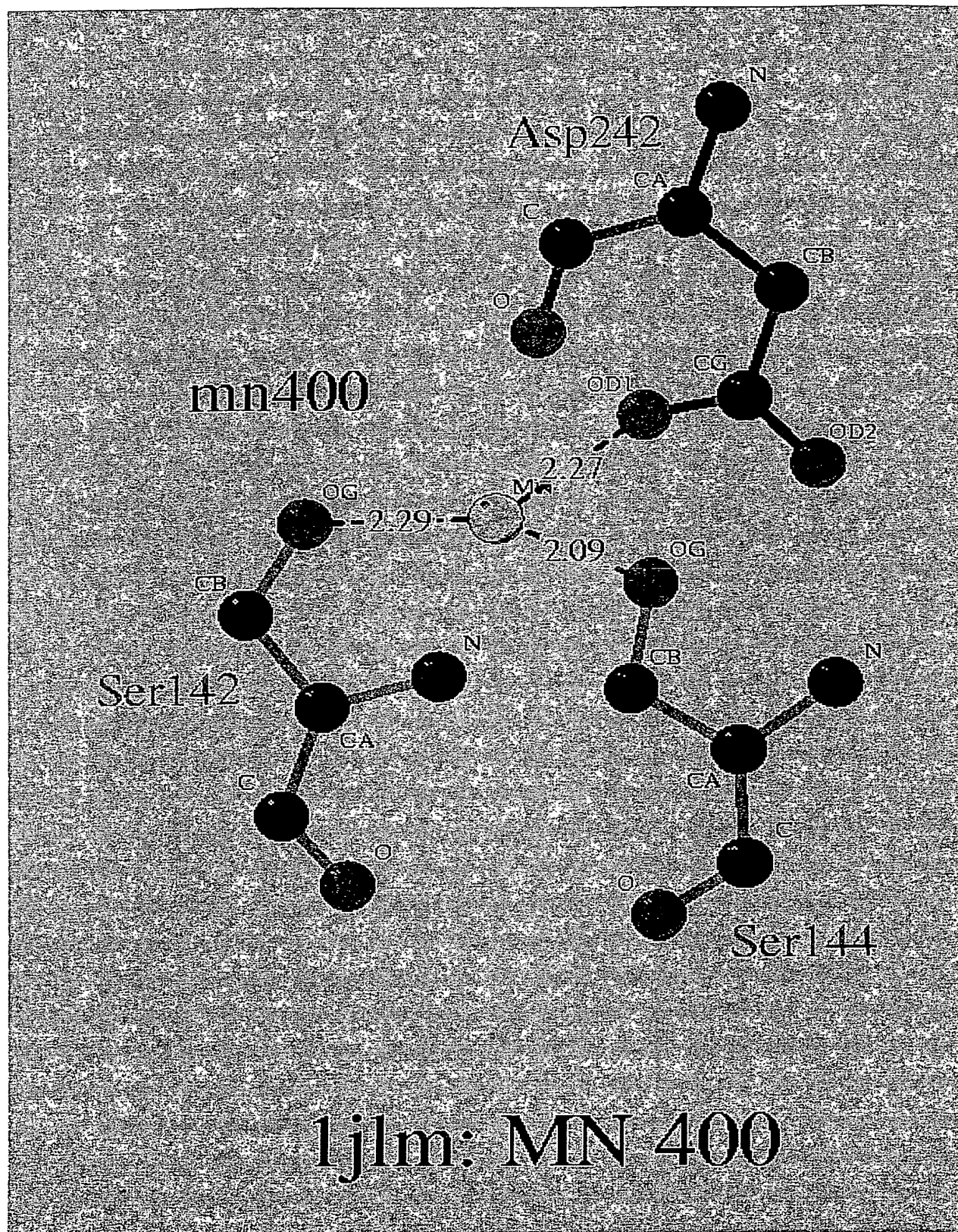
```

AAD05645.1 1JUN M I E I S E W L Q K L D D A L D K V V A K K E P E S F L K P I I S P I E D Y Q K S V R Q I Q A Q T D A P K F N E E G A Y P Q F L S C G L L Q V R G K N G
80 10 20 30 40 50 60 70
-----
AAD05645.1 1JUN A N M E F L L P K V Y P P P K S L Y I E H E K D G Q F L R E M L M P L L S S A P L V Q L E V I L I D A L S G G I F N L A P R L L D K N N D F I Y Q Q R
160 90 100 110 120 130 140 150
-----
AAD05645.1 1JUN I L T E S K E I E E A L K H L H E Y L K V N L Q E K L A G F R D F V H Y N E N A K D S L P L K A L F L S G V D A L S K D A L Y Y L E K I M R F G S K N G V
170 180 190 200 210 220 230
-----
AAD05645.1 1JUN L S F V N L E S E K N N Q S A E D L K R Y A E F F K D R T S F E C L K Y L N V E I I S D Q G I K S Q H M Q D F A D K I K A Y Y K Q K K E V K R E L K D L Q
240 250 260 270 280 290 300
-----
AAD05645.1 1JUN R D K E F W T K S S Q H E V S V P V G W D I N H K E V C F K I G N E Q N H T L I C D H S G S - G K S N F L H V L - - I Q N L A F Y Y D P D E V Q L F L L D
310 320 330 340 350 360 370 380
-----
AAD05645.1 1JUN Y K E G V E F N A Y V A D P A L E H A R L V S V A S S I S Y - - G I T F L K W L C D E M Q K R A D R F K Q F N V K D L S D Y R K H E K M P R L I V V I D E
390 400 410 420 430 440 450 460
-----
AAD05645.1 1JUN F Q V L F S D N K S T K A V E G H L N T L L K K G R S Y G V H L - V L A T Q T M R G T D I N P S F K A Q I A N R I A L P M D A E D S S S V L G D D A A C E
470 480 490 500 510 520 530
-----
AAD05645.1 1JUN I Q K P - - E G I F N N N G G N R K Y H T K M S V P K A P D D F K S F L T K I H A E F N Q R N L A P I D R K I Y N G E T P L K M P D T L K A N E M P L H L
540 550 560 570 580 590 600 610
-----
AAD05645.1 1JUN G K K V D Y E Q K D L I V E F E S N E S H L L V V I Q D L N A R I A L M K L L F Q N V K S A N K E L V F C N K E K R L I R S F D A Q K E Y G I T P V E N I
620 630 640 650 660 670 680 690
-----
AAD05645.1 1JUN L S V L D T A M N P N S A L V I D N L N E A K E L H D K V G A E K L K S F L E K A I D N E Q Y C V I F A H D F R Q I K T N Y H F D K L K E L L N N H F K Q
700 710 720 730 740 750 760 770
-----

```

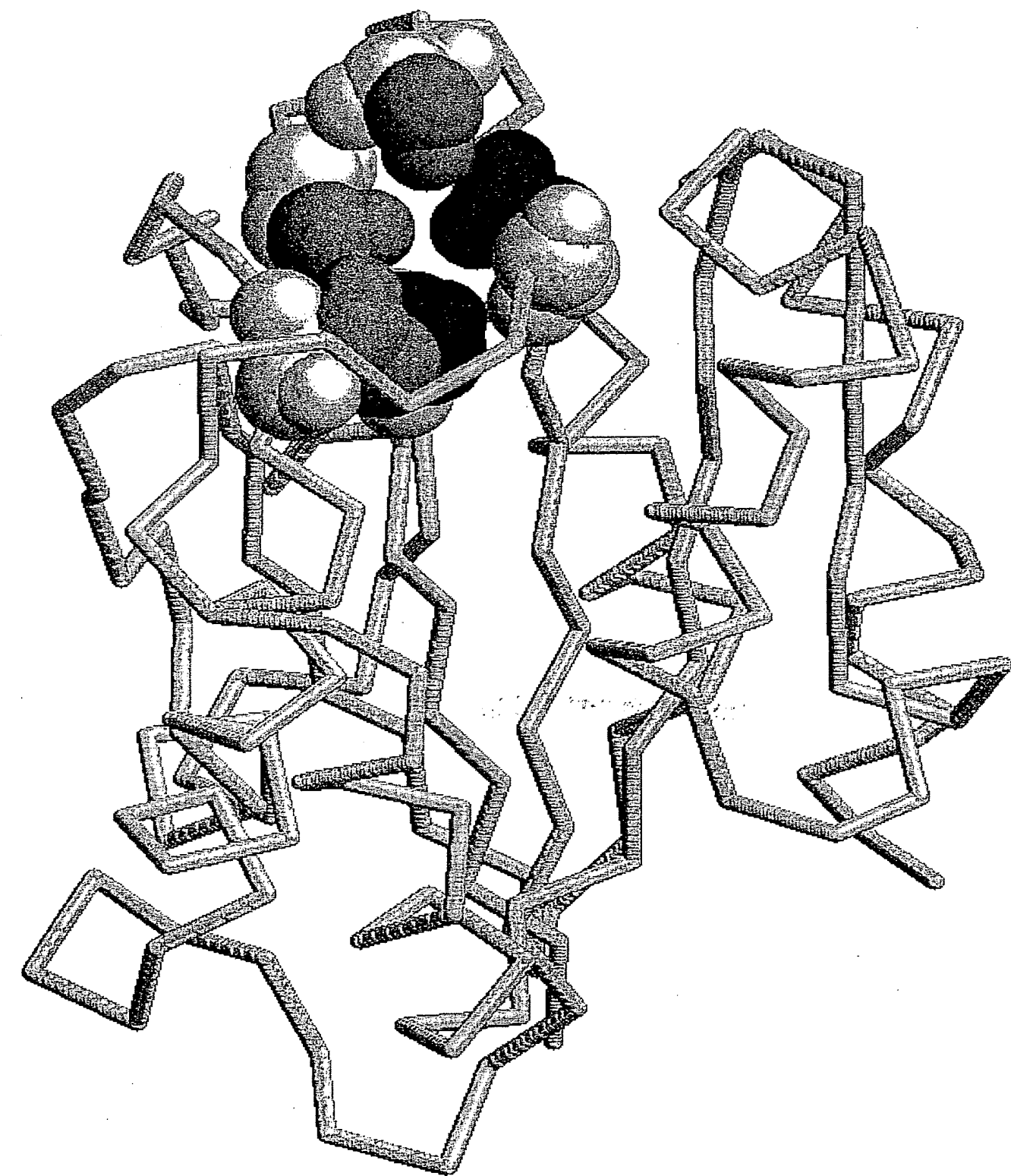
THIS PAGE BLANK (USPTO)

Figure 24A



THIS PAGE BLANK (USPTO)

● Figure 24B



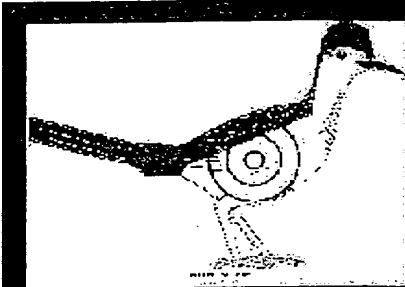
THIS PAGE BLANK (USPTO)

Figure 25


File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location <http://victoria.inpharmatica.co.uk/~volker/BPD3target.html>



Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):
- OR
- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

THIS PAGE BLANK (USPTO)

Figure 26A

2) 79 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st lter. (PSI)
<input checked="" type="checkbox"/>	P56199 drill through Top50BlastHits Red Seq. View	P56199	INTEGRIN ALPHA-1 (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD49A).	Homo sapiens (Human).	PRI	99%, 98% <u>unmasked SW</u>	1-192, 1-192	140-331, 140-331	488	100% <u>unmasked GT</u>	1
<input checked="" type="checkbox"/>	AAF01258.1 drill through Top50BlastHits Red Seq. View	AAF01258.1	integrin alpha-11 subunit precursor	Homo sapiens	PRI	57.7%, 57% <u>unmasked SW</u>	4-192, 2-192	163-349, 161-349	462	100% <u>unmasked GT</u>	1
<input checked="" type="checkbox"/>	AAD51919.2 drill through Top50BlastHits Red Seq. View	AAD51919.2	integrin alpha 11 subunit precursor	Homo sapiens	PRI	57.7%, 57% <u>unmasked SW</u>	4-192, 2-192	163-349, 161-349	462	100% <u>unmasked GT</u>	1
<input checked="" type="checkbox"/>	P17301 drill through Top50BlastHits Red Seq. View	P17301	PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GPIA) (COLLAGEN RECEPTOR) (INTEGRIN ALPHA-2) (VLA-2 ALPHA CHAIN) (CD49B).	Homo sapiens (Human).	PRI	52.9%, 52% <u>unmasked SW</u>	4-192, 4-192	173-361, 173-361	444	100% <u>unmasked GT</u>	1
<input checked="" type="checkbox"/>	Q99715 drill through Top50BlastHits Red Seq. View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	34.5%, 30% <u>unmasked SW</u>	1-192, 2-187	136-316, 2320-2495	442	100% <u>unmasked GT</u>	1
<input checked="" type="checkbox"/>	AAC31952.1 drill through Top50BlastHits Red Seq. View	AAC31952.1	integrin subunit alpha 10 precursor	Homo sapiens	PRI	47.6%, 47% <u>unmasked SW</u>	4-192, 2-192	166-354, 164-354	428	100% <u>unmasked GT</u>	1
<input checked="" type="checkbox"/>	AAA59544.1 drill through Top50BlastHits Red Seq. View	AAA59544.1	Not given	Homo sapiens	PRI	28.8%, 30% <u>unmasked SW</u>	5-192, 5-192	150-332, 150-332	421	100% <u>unmasked GT</u>	1
<input checked="" type="checkbox"/>	AAB24821.1 drill through Top50BlastHits Red Seq. View	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	28.8%, 30% <u>unmasked SW</u>	5-192, 5-192	150-332, 150-332	421	100% <u>unmasked GT</u>	1
<input checked="" type="checkbox"/>	AAB38702.1 drill through Top50BlastHits Red Seq. View	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	27.3%, 28% <u>unmasked SW</u>	1-192, 1-190	271-451, 37-218	421	100% <u>unmasked GT</u>	1
<input checked="" type="checkbox"/>	AAC01506.1 drill through Top50BlastHits Red Seq. View	AAC01506.1	type XII collagen	Homo sapiens	PRI	28.9%, 29% <u>unmasked SW</u>	1-192, 1-192	134-317, 134-317	412	100% <u>unmasked GT</u>	1

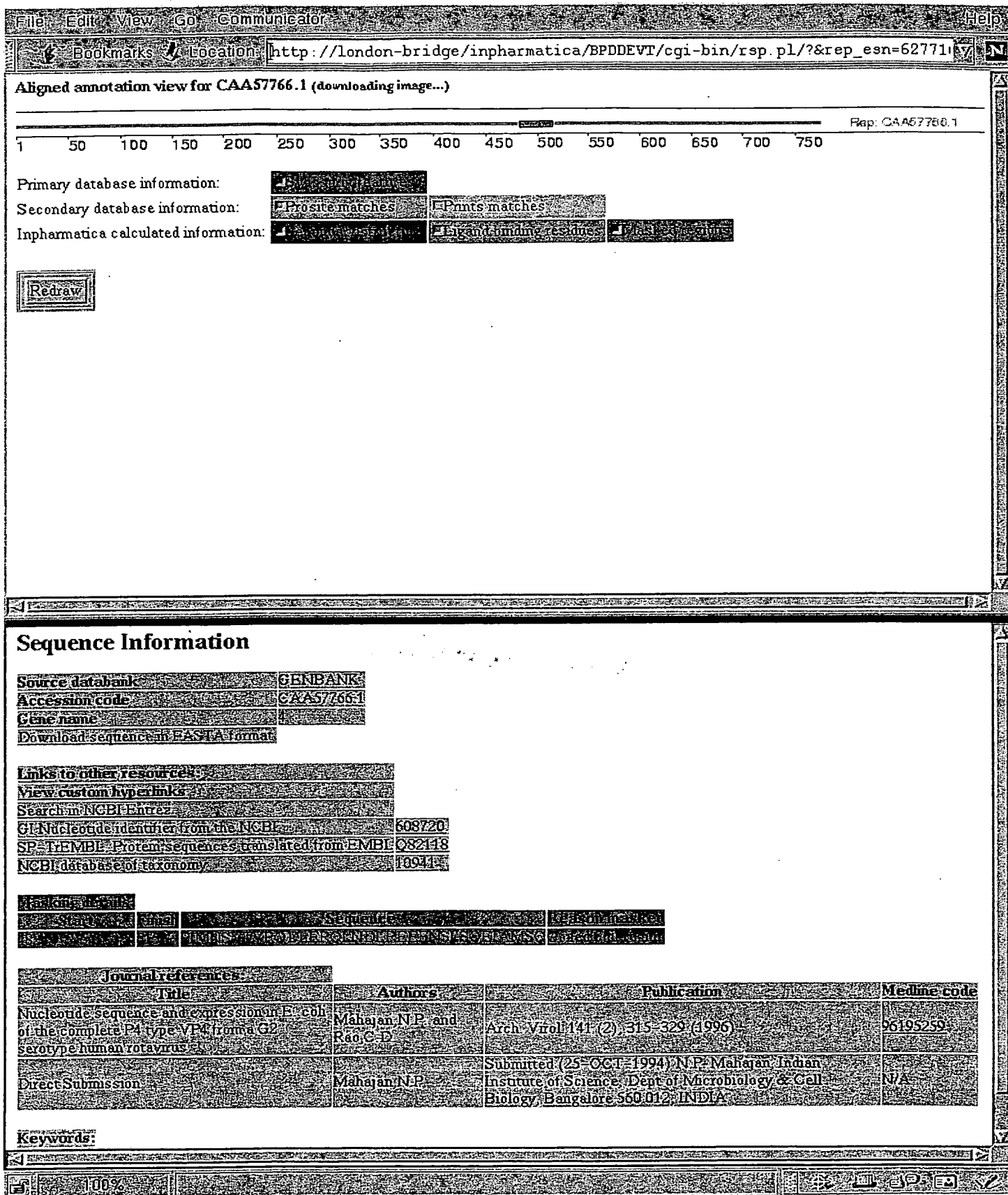
THIS PAGE BLANK (USPTO)

File Edit View Copy Communicator Help											
Back Forward Reload Home Search Navigate Print Security Shop											
Bookmarks Location http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl											
<input type="checkbox"/>	Red Seq. View	AAA47332.1 drill through Top50BlastHits	AAA47332.1	nonstructural protein S3	Human rotavirus	VRL	15.7% unmaskedSW	20-135	132-254	56	55.97% unmaskedGT
<input type="checkbox"/>	Red Seq. View	AA454009.1 drill through Top50BlastHits	AA454009.1	3-hydroxyacyl-CoA dehydrogenase, isoform 2	Homo sapiens	PRI	17.4% unmaskedSW	70-187	137-245	66	54.42% unmaskedGT reverse Hit
<input type="checkbox"/>	Red Seq. View	AAA47335.1 drill through Top50BlastHits	AAA47335.1	neutralizing protein VP4	Human rotavirus	VRL	14.3% unmaskedSW	82-189	631-770	69	53.19% unmaskedGT
<input type="checkbox"/>	Red Seq. View	P12474 drill through Top50BlastHits	P12474	OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4) [CONTAINS: OUTER CAPSID PROTEINS VPS AND VP8]	Bovine rotavirus (strain UK)	VRL	11.6% unmaskedSW	3-137	524-651	65	53.19% unmaskedGT
<input type="checkbox"/>	Red Seq. View	AA452707.1 drill through Top50BlastHits	AA452707.1	Not given	Homo sapiens	PRI	13.5% unmaskedSW	3-122	137-286	63	53.19% unmaskedGT
<input type="checkbox"/>	Red Seq. View	AA473046.1 drill through Top50BlastHits	AA473046.1	candidate tumor suppressor protein DICE1	Homo sapiens	PRI	11.5% unmaskedSW	6-47	4-55	72	51.98% unmaskedGT
<input type="checkbox"/>	Red Seq. View	CAA65528.1 drill through Top50BlastHits	CAA65528.1	3-hydroxyacyl-CoA dehydrogenase	Homo sapiens	PRI	16.5% unmaskedSW	70-187	78-186	65	51.98% unmaskedGT reverse Hit
<input type="checkbox"/>	Red Seq. View	AAD50337.1 drill through Top50BlastHits	AAD50337.1	terminal protein	fowl adenovirus 8	VRL	15.5% unmaskedSW	9-131	93-202	59	49.61% unmaskedGT
<input type="checkbox"/>	Red Seq. View	AAA50482.1 drill through Top50BlastHits	AAA50482.1	VP4	Murine rotavirus	VRL	14.1% unmaskedSW	4-188	517-707	58	49.61% unmaskedGT
<input type="checkbox"/>	Red Seq. View	CAA57766.1 drill through Top50BlastHits	CAA57766.1	VP4	Human rotavirus	VRL	15.2% unmaskedSW	11-189	592-770	69	48.45% unmaskedGT
<input type="checkbox"/>	Red Seq. View	CAA28420.1 drill through Top50BlastHits	CAA28420.1	apo-B100 precursor	Homo sapiens	PRI	11.5% unmaskedSW	8-192	305-507	49	48.45% unmaskedGT
<input type="checkbox"/>	Red Seq. View	CAA28191.1 drill through Top50BlastHits	CAA28191.1	Not given	Homo sapiens	PRI	11.5% unmaskedSW	8-192	308-510	49	48.45% unmaskedGT
<input type="checkbox"/>	Red Seq. View	AAB04636.1 drill through Top50BlastHits	AAB04636.1	Not given	Homo sapiens	PRI	11.5% unmaskedSW	8-192	308-510	49	48.45% unmaskedGT
<input type="checkbox"/>	Red Seq. View	AAB00481.1 drill through Top50BlastHits	AAB00481.1	Not given	Homo sapiens	PRI	11.5% unmaskedSW	8-192	308-510	49	48.45% unmaskedGT
<input type="checkbox"/>	Red Seq. View	BAA02665.1 drill through Top50BlastHits	BAA02665.1	VP4	Canine rotavirus	VRL	14% unmaskedSW	4-192	518-712	61	46.19% unmaskedGT
<input type="checkbox"/>	Red Seq. View	BAA77540.1 drill through Top50BlastHits	BAA77540.1	VP4	Human rotavirus 3	VRL	15% unmaskedSW	62-189	631-770	65	45.1% unmaskedGT
<input type="checkbox"/>	Red Seq. View	AAA47333.1 drill through Top50BlastHits	AAA47333.1	Not given	Human rotavirus	VRL	14.3% unmaskedSW	13-189	594-770	63	45.1% unmaskedGT
<input type="checkbox"/>	Red Seq. View	AAA35549.1 drill through Top50BlastHits	AAA35549.1	Not given	Homo sapiens	PRI	11% unmaskedSW	8-192	308-510	49	45.1% unmaskedGT

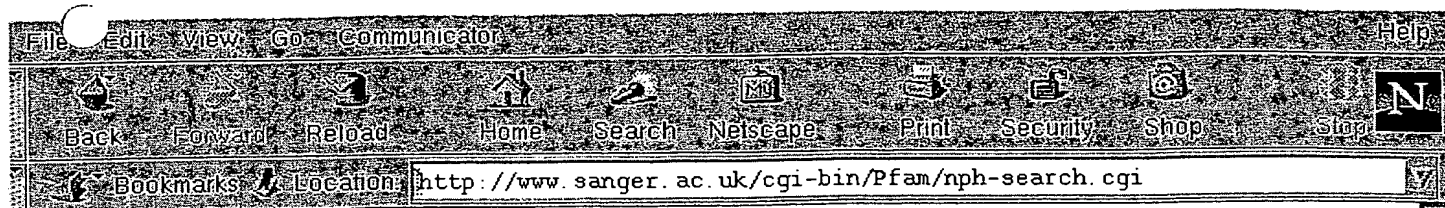
THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

Figure 27



THIS PAGE BLANK (USPTO)



Pfam

Protein families database of alignments and HMMs

[Home](#) | [Keyword search](#) | [Protein search](#) | [DNA search](#) | [Browse Pfam](#) | [Taxonomy search](#) | [Help](#)



Results for gi|608720|emb|CAA57766.1|

Trusted matches - domains scoring higher than the gathering threshold

Domain	Start	End	Bits	Evalue	Alignment
VP4	1	775	1854.50	0	Align

[775 residues]

VP41-775

Alignments of Pfam-A domains to HMMs

Format for fetching alignments to seed

[Hypertextlinked to swisspfam](#)

Alignment of VP4 vs gi|608720|emb|CAA57766.1|1-775:

```

*->MASLIYRQLLSNSYTVdldeindigsektqngvtnpGPFaqtgya
gi|608720| 1 MASLIYRQLL+nSY+vdl dei++igsektqn vt+npGPFaqt+ya 46
MASLIYRQLLTNSYSVDLYDEIEIQIGSEKTQN-VTINPGPFAQTRYA

PvkwhggetndsttvepvldePYtntliqpdsfndpvqywmLlnptndGv
gi|608720| 47 PvkwhggetndsttvepvldePY qp++f++p++yw L++++dGv 51
PVDWGHGEINDSTTVEPVLDEPY-----QPTTFKPPNDYWLIISSNTDGV

VvegTnnsdrwiayvPvltqtsrqYtlfgqnkqisVdntsdtkwkFv
gi|608720| 92 VYESTNNSDFWTAIVAEVPHVSQTHRQYVLFGENKQFNUENNS-DKWKFF 140
V++Tnnsd++a+++veP+v+q+rqY+lfq+nkq++v+n+s +kwkF+

evikttdlgnvysrslslnklagimkkgvgrlygyrgetpnattguyt
gi|608720| 141 EMPKGSQSDFSNRRLTSTNNRLVGMMLKYG-GRVWTFHGETPRATDSSN 189
+++ +++++r+tl+s+n+l g++k+g gr++++getp+att++++

vsnefpniqvnlnvnyfiiprsqeskc+eyinnGLPpiQntnrvpvsis
gi|608720| 190 TAB-LNNISIIITHSEFYIIPRSQESKCEYINNGLPPIQNTNRNVVPLSL 238
+++ ++ni+++i+++fYiiprsqeskc+eyinnGLPpiQntnrvp+s+s

srsIye.raqpNeDiVISKtSLWKEvqYnRDIiirfkFansiikSgGGLGY
gi|608720| 239 srsI+++raq+NeDi+ISKtSLWKE+qYnRDIiirfkF+ns+ik+GGLGY 288
SRSIQYrRAQVNHEDITI SKTSLWKEMQYNRDIIRFKFGNSVIKLGGLGY

kwsveisFkaanyqYtytRDGeevNAHTTcSVNGvndfsyNgGSLPATDF
gi|608720| 289 KWS-EISYKAANYQYSYSRDGEQVTAHTTCSVNGVNNFSYNGGSLP-TDF 336
kws eiS+KaanyqY+Y RDGe+v+AMTTCsvNgv+n+fsyNgGSLP TDF

nisryevikensyvvYvdyWDDsqafRNMVYVRSLaAnsrlngvltcg6sY
gi|608720| 337 SISRVEVIKGINVYVIDYWDDSKAFRNMVYVRSLaAN ln+v+c+g6sY 384
SISRVEVIKGINVYVIDYWDDSKAFRNMVYVRSLaAN--LNSVKCVG6sY

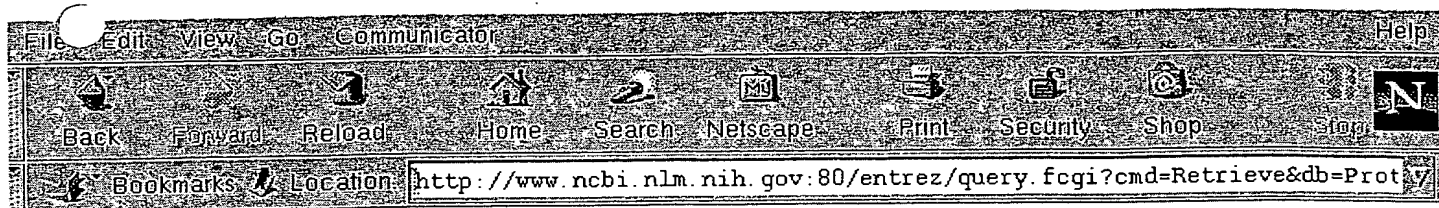
sFaLNYIPalRevrNKGKvNwPvmsGGaVqLhsagvTLStqFTDFVSLN
gi|608720| 385 +F+L P V G wP+m+GGaV+Lh+agvTLStqFTDFVSLN 421
DFRL---P---V---G---EWPIMNGGAVSLHFAGVTLSTqFTDFVSLN

SLRFRddFrlavdnklePfFsiirtrvtrSGGdlYrqGLPAAnPngYdde
gi|608720| 422 SLRFR-F+L+Vd eP+Psiirtr+ + LY GLPAAnPn+ ++e 459
SLRFR--FSLTVD---EPSPsiirtrvtrSGGdlYrqGLPAAnPn+ ++e

yyEiaGrfSLISLVP+ DDYQTPI+N SVTVRqDLERqLndLReeFNSLS
yyE++GrfSLISLVP+ DDYQTPI+N SVTVRqDLERqLndLReeFNSLS

```

THIS PAGE BLANK (USPTO)



NCBI Entrez Protein

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM

Search: Protein for

as

☐ 1: GI "608720" [GenPept]

VP4 [Human rotavirus]

[BLink](#), [PubMed](#), [Related Sequences](#), [Nucleotide](#), [Taxonomy](#)

LOCUS CAA57766 775 aa VRL 06-AUG-1996

DEFINITION VP4 [Human rotavirus].

ACCESSION CAA57766

PID g608720

VERSION CAA57766.1 GI:608720

DBSOURCE emb1 locus HRVP4, accession X82323.1

KEYWORDS

SOURCE Human rotavirus.

ORGANISM Human rotavirus

REFERENCE 1 (residues 1 to 775)

AUTHORS Mahajan, N.P. and Rao, C.D.

TITLE Nucleotide sequence and expression in E. coli of the complete P4 type VP4 from a G2 serotype human rotavirus

JOURNAL Arch. Virol. 141 (2), 315-329 (1996)

MEDLINE 36195259

REFERENCE 2 (residues 1 to 775)

AUTHORS Mahajan, N.P.

TITLE Direct Submission

JOURNAL Submitted (25-OCT-1994) N.P. Mahajan, Indian Institute of Science, Dept of Microbiology & Cell Biology, Bangalore 560 012, INDIA

FEATURES

source Location/Qualifiers

1..775

/organism="Human rotavirus"

/strain="IS-2"

/isolate="G serotype 2"

/db_xref="taxon:10941"

Protein 1..775

/product="VP4"

CDS 1..775

/gene="4"

/db_xref="SPTREMBL:Q82118"

/coded_by="608719:10..2337"

ORIGIN

1 masliyrqll tnsysvdlpd eiegigsekt qnvtinpgpf aqtryapvdw ghgeindstt

61 vepvldgpyq ptffkppndy wllissntdg vvestnnsd fwtaviavep hvsqtarqyv

121 lfgenkqfnv ennsdkwkff emfkgssqsd fsnrrtltsn nrlvgmlkyg grwvtfhget

181 prattssnt adlnnisiii hsefyiiprs qeskeneyin nglppiqntr nvvpplsissr

241 siqyrraqm editisktsl wkemqynrdi iirfkfgmsv iklgglgykw seisykaany

301 qysysrdgeq vtahttcsvn gvnnfsyngg slptdfsisr yevikgihyv yidywddska

361 frnmvyvrsi aanlnsvkcv ggsydfripv gewpimngga vslhifagvtl stqftdfvsl

421 nslrfrfslt vdepsfsiir trtnnlyglp aanpnngney pevsggrfsli slvptiddyq

481 tpimnsvtvr qdlerqlndi reefnslsqe iamsqlidla llpldmfsmf sgikstidlt

541 ksmatsvmkk frksklatsi semtnslsda assasrsvsi rsnfstisnw sdasksvlnv

601 tdsvndvstg tsiiskklrl remitqtegi sfddisaavl ktkidmstgi gkntlpdvvt

661 edsekfipkr syrvlkdnev meintegkff aykvdtlnei pfdinkfael vtdspvisai

721 idfkrklknln dnygitriea lnliksnpnv lnmfingmp iimrieqli lqckl


//

THIS PAGE BLANK (USPTO)

● Figure 30A

File Edit Search Type Help

Accession Code Query



Tools

←

→

?

?

?

X

Sequences

?

?

?

? Acc
y122

Ligands

? ADP
NAD
FAD

? Phe
Ala
Glu

Select a database and enter a code

Database	
GenBank	
Accession	<input type="radio"/>
Accession Version	<input type="radio"/>
GI-NID	<input type="radio"/>
protein_id	<input checked="" type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

caa57766.1

Submit Reset

THIS PAGE BLANK (USPTO)

Figure 30B

File Edit Search Type Help

Aligned Sequence Display

Query details:
CAA57766.1: VP4

inpharmatica

inpharmatica

Tools

Sequences

Legends

Page 1 of 1

Total hits: 11

Selected: 0

Cluster Details

Code

Title

Organism

%ID

Query origin

Target origin

Align. score

Method

Confidence

1

1ATZB

HUMAN VON WILLEBRAND FACTOR A3 DO...

Hsapiens

12.3

523 - 722

5 - 185

57

Local

Low(50%)

2

1QC5A

DOMAIN FROM INTEGRIN ALPHA1 - BETAI...

N/A

15.2

592 - 770

11 - 189

69

Local

Low(48%)

3

1C9NA

BACILLUS LENTUS SUBSTITISIN VARIANT ...

Blentus

12.5

317 - 480

86 - 239

65

Local

Low(42%)

4

1AX4B

TRYPTOPHANASE FROM PROTEUS VULGA...

P.vulgaris

21.4

619 - 745

170 - 288

84

Local

Low(33%)

5

1AX4C

TRYPTOPHANASE FROM PROTEUS VULGA...

P.vulgaris

21.4

619 - 745

170 - 288

84

Local

Low(33%)

6

1AX4D

TRYPTOPHANASE FROM PROTEUS VULGA...

P.vulgaris

21.4

619 - 745

170 - 288

84

Local

Low(33%)

7

1ESF-A

STAPHYLOCOCCAL ENTEROTOXIN A

Saureus

17.8

250 - 355

43 - 137

66

Local

Low(21%)

8

1AX4A

TRYPTOPHANASE FROM PROTEUS VULGA...

P.vulgaris

21.4

619 - 745

170 - 287

74

Local

Low(18%)

9

1CJA-A

ACTIN- FRAGMIN KINASE, CATALYTIC DO...

Ppolicephalum

21.9

643 - 771

33 - 154

68

Local

Low(14%)

10

1CJA-B

ACTIN- FRAGMIN KINASE, CATALYTIC DO...

Ppolicephalum

21.9

643 - 771

33 - 154

68

Local

Low(14%)

11

1AOMA

SUBSTRATE AND PRODUCT BOUND TO CY...

Pdenitrificans

10.2

47 - 393

6 - 367

44

Local

Low(13%)

inpharmatica

inpharmatica

GenBank

SwissProt

PDB

THIS PAGE BLANK (USPTO)

Aligned Sequence Display

Query details:
CAA57786.1:VP4

Cluster	Detail	Code	Title	%ID	Query idn.	Target idn.	Bitter.	E value
Δ	AAA18376.1		VP4 protein	95.0	1 - 775	1 - 775	1	0.00E+00
Δ	AAA47333.1		Not given	94.0	1 - 775	1 - 775	1	0.00E+00
Δ	AAA47335.1		neutralizing protein VP4	94.0	1 - 775	1 - 775	1	0.00E+00
Δ	P111196		OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OU)	94.0	1 - 775	1 - 775	1	0.00E+00
Δ	CAB82923.1		outer capsid protein 4	93.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77540.1		VP4	92.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77541.1		VP4	89.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77549.1		VP4	89.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77542.1		VP4	89.0	1 - 775	1 - 775	1	0.00E+00
Δ	AAB0555		outer capsid VP4	89.0	1 - 775	1 - 775	1	0.00E+00
Δ	P13842		OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OU)	89.0	1 - 775	1 - 775	1	0.00E+00
Δ	AAA114298.1		VP4 protein	89.0	9 - 286	1 - 278	1	0.00E+00
Δ	BAA77550.1		VP4	89.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77556.1		VP4	89.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77555.1		VP4	89.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77538.1		VP4	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77539.1		VP4	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77544.1		VP4	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77546.1		VP4	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	P111195		OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OU)	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	AAD50894.1		VP4	88.0	1 - 286	1 - 286	1	0.00E+00
Δ	AAD50892.1		VP4	88.0	1 - 286	1 - 286	1	0.00E+00
Δ	BAA03856.1		VP5 and VP8	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	AAA47290.1		Not given	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77553.1		VP4	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77551.1		VP4	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77548.1		VP4	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77547.1		VP4	88.0	1 - 775	1 - 775	1	0.00E+00
Δ	BAA77545.1		VP4	87.0	1 - 775	1 - 775	1	0.00E+00
Δ	AAAG6953.1		outer capsid protein	87.0	1 - 775	1 - 775	1	0.00E+00

Total selected: 0 View alignment

Reverse Maximised PSI-BLAST results.

Page 1 of 6 Total hits: 168 Selected: 0

Select all Deselect all

Sequences

Ligands

D GenBank D SWISS-PROT D PDB

THIS PAGE BLANK (USPTO)

ALEye output (January 2, 2001 12:53 PM)

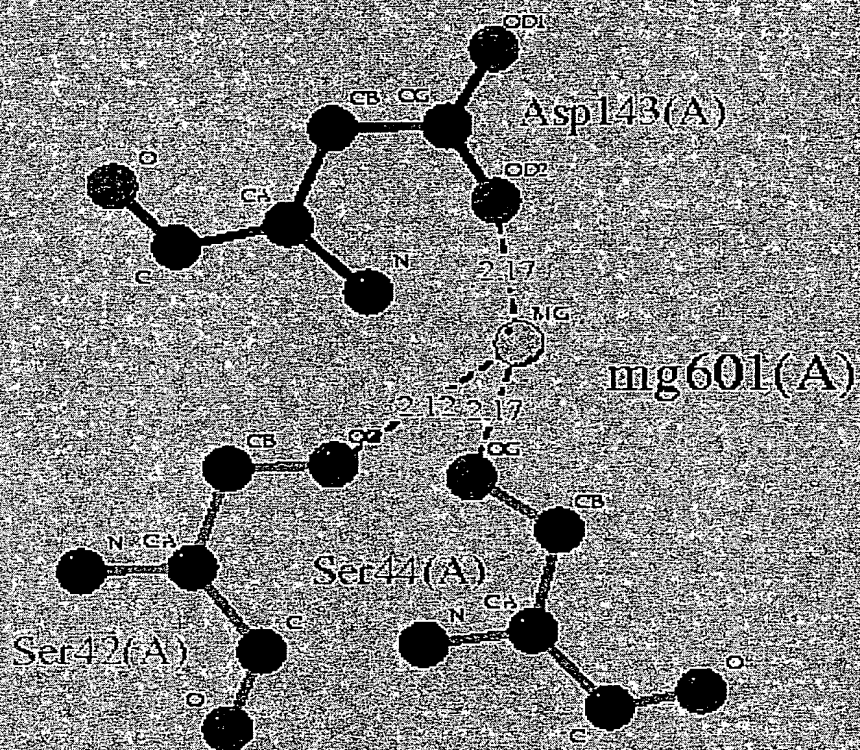
```

CAA57788.1      10      20      30      40      50      60      70
1QCS:A  MASLIYRQLLTNSYSVDLYDEIEQIGSEKTNVTINPGPFAQTRYAPVDWGHGEINDSTTVEPVLDGPYQPTTFKPP
      80      90      100      110      120      130      140      150
CAA57788.1  NDYWLLISSNTDGVVYESTNNSDFWTAVIAVEPHVSQTNRQYVLFGENKQFNVENNSDKWKFFEMFKGSSQSDFSNR
1QCS:A      160      170      180      190      200      210      220      230
      RTLTNNRLVGLKYGGRVWTFHGETPRATTDSSNTADLNNISIIHSEFYIIPRSQESKCNEYINNGLPPIQNTRN
CAA57788.1
1QCS:A
      240      250      260      270      280      290      300
      VVPLSLSSRSIQYRRAQVNEEDITISKTSLWKEMQYNRDIIRFKFGNSVIKLGGLGYSYKAANYQYSYSPDG
CAA57788.1      310      320      330      340      350      360      370      380
1QCS:A  EQVTAHTTCSVNGVNNFSYNGGSLPTDFSISRVEYIKGIHYVYIDYWDSDSKAFNMVYVVRSLAANLNSVKCVGGSYD
      390      400      410      420      430      440      450      460
CAA57788.1  FRLPVGEWPIMNGGAVSLHFAGVTLSTQFTDFVSLNSLRRFRSLTVDEPSFSIIRTRTMNLYGLPAANPNNGNEYE
1QCS:A
      470      480      490      500      510      520      530
      VSGRFLSLVPTIDDYQTPIMNSVTVRQDLERQLNDLPEEFNSLSQEIAMSQDLIDLALLPLDMFSMFGIKSTIDL
CAA57788.1      540      550      560      570      580      590      600      610
1QCS:A  TKSMATSVMKKFRKSKLATSISEMNTSLSDAASSASRVSIRSNFSTISNWS-----DASKSVL---NVTDSV
      620      630      640      650      660      670      680      690
      NDVSTQTSII SKKLRLP-----EMITQTEGISFDDISAAVLTKIDMSTQIGKNTLPDVVTTEDSEK---FIPKR---SYR
CAA57788.1      700      710      720      730      740      750      760      770
1QCS:A  NDLLERMDIGPKQTQVGIVQYGENVTHEFNLNKYSSTEEVLVAAKKIVQRRGRQMTALGTDTPRKEAFTEARPGAPR
      780      790      800      810      820      830      840      850
      VLNKNEVMEINTEGKFFAYKYVDTLNEIPFDINKFAELVTDSPVISAIDFKRLKNLNDNYGITRIEALNLKSNPNV
CAA57788.1      860      870      880      890      900      910      920      930
1QCS:A  GYK--KVMVIVTDGE--SHDNHRLKKVIQDCED--ENIQRFSI--AILGSYNPGNLSTKEKFVEEIKS---IASSEP-T

```

THIS PAGE BLANK (USPTO)

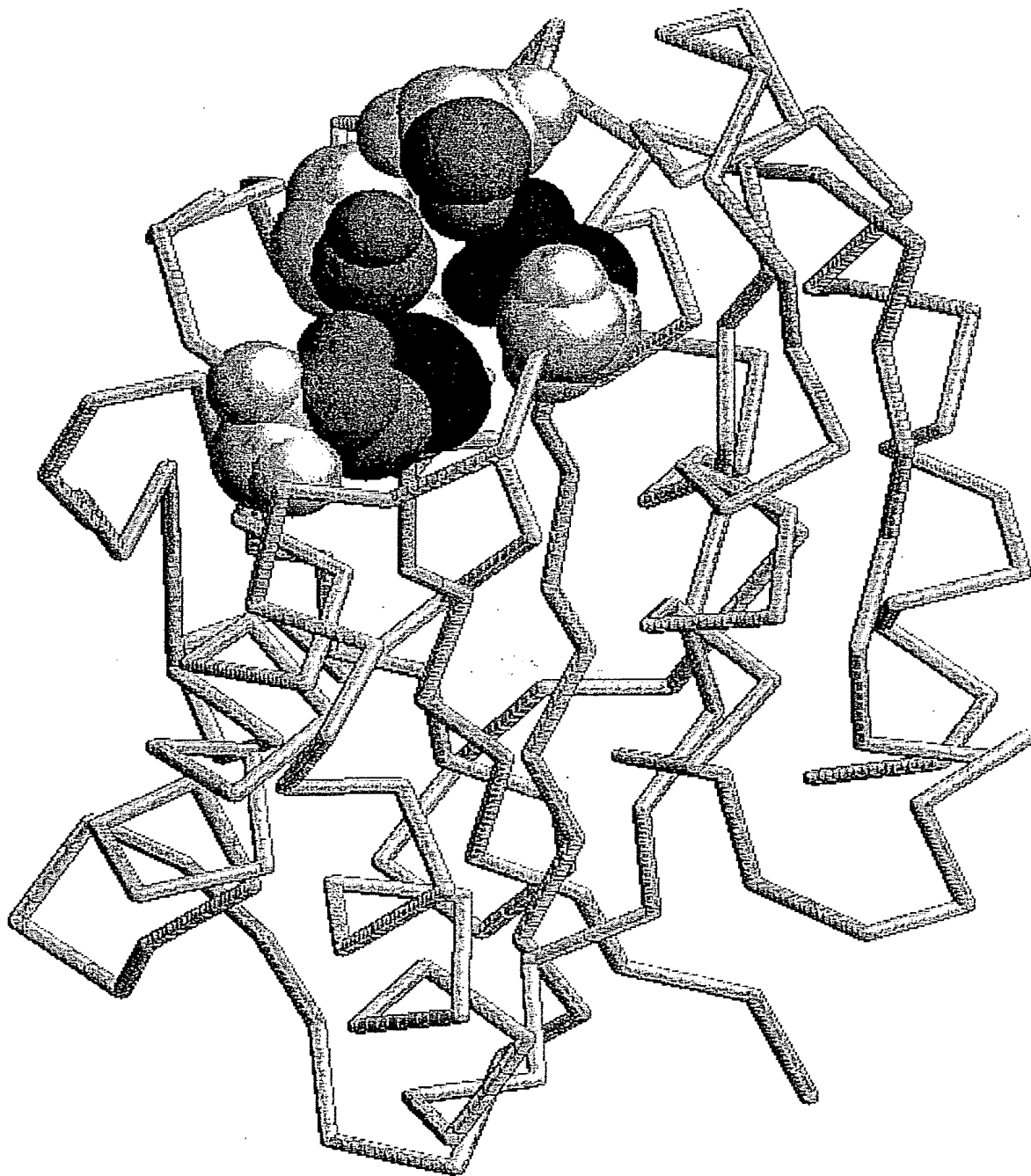
Figure 32A



1qc5: MG 601 Chain [A]

THIS PAGE BLANK (USPTO)

● Figure 32B

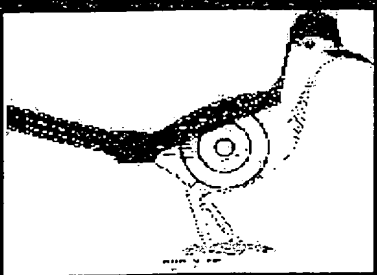


THIS PAGE BLANK (USPTO)


Figure 33

FileEditViewGoCommunicatorHelp

BookmarksLocationhttp://victoria.inpharmatica.co.uk/~volker/BPD3target.html



Target Mining Interface



Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):
- OR
- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters


- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:

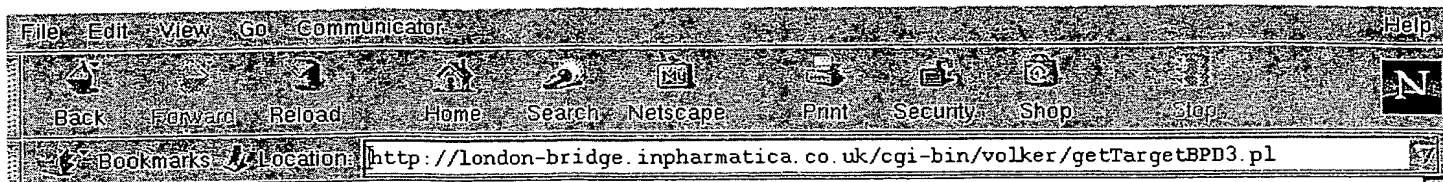
☐ Homo sapiens☐ Rattus norvegicus (Rat)☐ Mus musculus (Mouse)☐ Danio rerio (Zebrafish)

100%



THIS PAGE BLANK (USPTO)

Figure 34A



2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI - Blast output: PSI - BLAST values are shown in maroon!

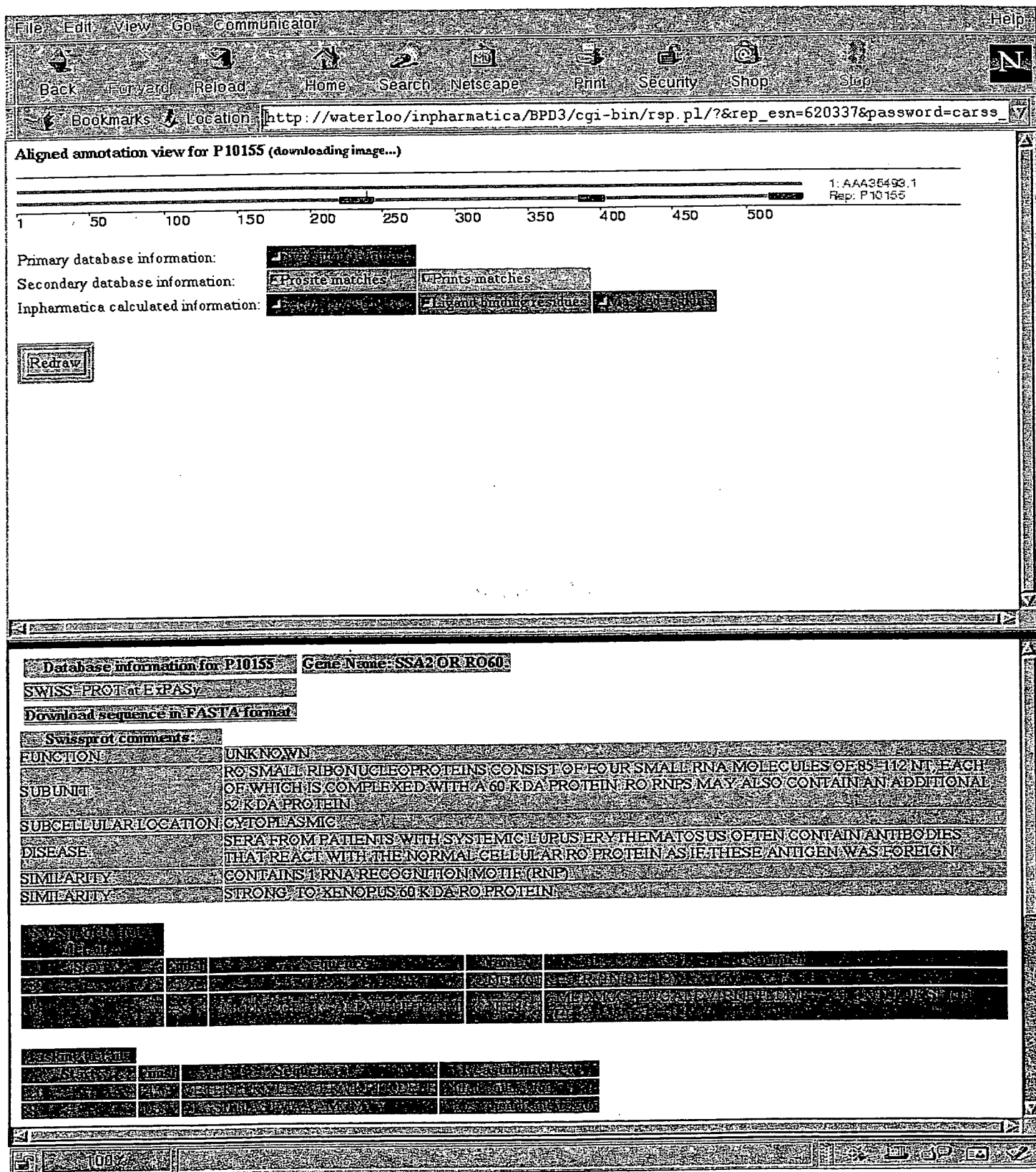
Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rqn. (GT,PSI)	Target rqn. (GT,PSI)	Aln. score (GT)	Conf. (GT)	1st lter. (PSI)
<input type="checkbox"/>	AAA59544.1 drill through Top50BlastHits Red Seq. View	AAA59544.1	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
<input checked="" type="checkbox"/>	AAB24821.1 drill through Top50BlastHits Red Seq. View	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
<input type="checkbox"/>	Q99715 drill through Top50BlastHits Red Seq. View	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.9%, 28% unmaskedSW	2-186, 2-179	439-617, 2322-2494	456	100% unmaskedGT	1
<input type="checkbox"/>	AAB38702.1 drill through Top50BlastHits Red Seq. View	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	28.9%, 25% unmaskedSW	2-186, 2-166	274-452, 40-221	446	100% unmaskedGT	1
<input type="checkbox"/>	AAC01506.1 drill through Top50BlastHits Red Seq. View	AAC01506.1	type XII collagen	Homo sapiens	PRI	28.4%, 28% unmaskedSW	2-186, 2-186	137-318, 137-318	445	100% unmaskedGT	1
<input type="checkbox"/>	CAA72402.1 drill through Top50BlastHits Red Seq. View	CAA72402.1	collagen type XIV	Homo sapiens	PRI	28.7%, 30% unmaskedSW	2-186, 2-186	6-185, 6-185	442	100% unmaskedGT	1
<input type="checkbox"/>	AAB38547.1 drill through Top50BlastHits Red Seq. View	AAB38547.1	leukointegrin alpha d chain	Homo sapiens	PRI	61%, 60% unmaskedSW	1-187, 1-187	148-334, 148-334	439	100% unmaskedGT	1
<input type="checkbox"/>	CAB71222.1 drill through Top50BlastHits Red Seq. View	CAB71222.1	dJ238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	27.1%, 22% unmaskedSW	1-186, 2-186	293-472, 1430-1620	439	100% unmaskedGT	1
<input type="checkbox"/>	CAA07569.1 drill through Top50BlastHits Red Seq. View	CAA07569.1	matrilin-4	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-187	344-525, 34-215	418	100% unmaskedGT	2
<input type="checkbox"/>	CAB46380.1 drill through Top50BlastHits Red Seq. View	CAB46380.1	dJ453C12.3 (matrilin-4)	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-186	385-566, 34-214	418	100% unmaskedGT	2

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

Figure 35



THIS PAGE BLANK (USPTO)

Figure 36

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Bookmarks Location: <http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi>

The Sanger Centre **Pfam** Protein families database of alignments and HMMs

Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help

Results for gi|133251|sp|P10155|RO60_HUMAN

There were no matches to Pfam-A (including borderline matches) for gi|133251|sp|P10155|RO60_HUMAN

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_8344	1	194	2.3e-103	Align
Pfam-B_10162	195	538	1.8e-165	Align

[538 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: [Hypertext linked to swisspfam](#)

Query gi|133251|sp|P10155|RO60_HUMAN/1-194 matching Pfam-B_8344

```

Q92787 1 MEE SVNQM QPLNEKQ IAH S QDGYVWQVTD MNR LHRFLCF6 SEGG TTYIKE 50
gi|133251|sp|P10155|RO60_HUMAN 1 MEE SVNQM QPLNEKQ IAH S QDGYVWQVTD MNR LHRFLCF6 SEGG TTYIKE 50

Q92787 51 QKL GLENAEALIRLIEDGRGCEVIQEIKSF SQEGRTTKQEPMLFALAI CS 100
gi|133251|sp|P10155|RO60_HUMAN 51 QKL GLENAEALIRLIEDGRGCEVIQEIKSF SQEGRTTKQEPMLFALAI CS 100

Q92787 101 QCS DI STKQAAPKAVSEVCRIP THLFTFIQFKKDLKESMKCGMWGRALRK 150
gi|133251|sp|P10155|RO60_HUMAN 101 QCS DI STKQAAPKAVSEVCRIP THLFTFIQFKKDLKESMKCGMWGRALRK 150

Q92787 151 AIADWYNEKGGMALALAVTKYKQRNGWSHKDLLRLSHLKP SSEG 194
gi|133251|sp|P10155|RO60_HUMAN 151 AIADWYNEKGGMALALAVTKYKQRNGWSHKDLLRLSHLKP SSEG 194

```

[Align to family](#)

Query gi|133251|sp|P10155|RO60_HUMAN/195-538 matching Pfam-B_10162

```

008848 195 LAIVTKYITKGWKEVHEEYKEKAL SVEAEKLLKYLEAVEKVKRTKD DLEV 244
gi|133251|sp|P10155|RO60_HUMAN 195 LAIVTKYITKGWKEVHEEYKEKAL SVEAEKLLKYLEAVEKVKRTKD+LEV 244

008848 245 IHLIEEHQLVREHLLTNHLKSKEVWKALLQEMPLTALLRNLGKMTANSVL 294
gi|133251|sp|P10155|RO60_HUMAN 245 IHLIEEH+LVREHLLTNHLKSKEVWKALLQEMPLTALLRNLGKMTANSVL 294

008848 295 EPGNSEVSLICEKL SNEKLLKKARINPFHVLIALETYRAGHGLRGK LKWI 344
gi|133251|sp|P10155|RO60_HUMAN 295 EPGNSEVSL+CEKL NEKLLKKARINPFH+LIALETY+ GHGLRGK LKW 344

008848 345 PDKDILQALDRAFYTTFTKTEPTGKRFLLAUDV SASMNRALGSVLNAST 394

```

THIS PAGE BLANK (USPTO)

37

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop

Bookmarks Location <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=Prot>

LOCUS R060_HUMAN 538 aa PRI 01-FEB-1996
DEFINITION 60 KD RO PROTEIN (60 KD RIBONUCLEOPROTEIN RO) (RORNP) (SJ06REN SYNDROME TYPE A ANTIGEN (SS-A)).
ACCESSION P10155
PID g133251
VERSION P10155 GI:133251
DBSOURCE swissprot: locus R060_HUMAN, accession P10155;
class: standard.
created: Mar 1, 1989.
sequence updated: Mar 1, 1989.
annotation updated: Feb 1, 1996.
xrefs: gi: 177782, gi: gi: 177783, gi: gi: 387656, gi: gi: 387657, gi: gi: 86722, gi: gi: 107626
xrefs (non-sequence databases): MIM 600063, MIM 234700, PROSITE P500030
KEYWORDS Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus; Antigen.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 538)
AUTHORS Deutscher, S.L., Harley, J.B. and Keene, J.D.
TITLE Molecular analysis of the 60-kDa human Ro ribonucleoprotein
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9479-9483 (1988)
MEDLINE 89071722
REMARK SEQUENCE FROM N.A.
REFERENCE 2 (residues 1 to 538)
AUTHORS Ben-Chetrit, E., Gandy, B.J., Tan, E.M. and Sullivan, K.F.
TITLE Isolation and characterization of a cDNA clone encoding the 60-kD component of the human SS-A/Ro ribonucleoprotein autoantigen
JOURNAL J. Clin. Invest. 83 (4), 1284-1292 (1989)
MEDLINE 89198084
REMARK SEQUENCE FROM N.A.
COMMENT

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. The original entry is available from <http://www.expasy.ch/sprot> and <http://www.ebi.ac.uk/sprot>

[FUNCTION] UNKNOWN.
[SUBUNIT] RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT, EACH OF WHICH IS COMPLEXED WITH A 60 KD PROTEIN. RO RNPS MAY ALSO CONTAIN AN ADDITIONAL 52 KD PROTEIN.
[SUBCELLULAR LOCATION] CYTOPLASMIC.
[DISEASE] SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR RO PROTEIN AS IF THESE ANTIGEN WAS FOREIGN.
[SIMILARITY] CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
[SIMILARITY] STRONG, TO XENOPUS 60 KD RO PROTEIN.

FEATURES
source 1..538
/organism="Homo sapiens"
/db_xref="taxon:9606"
Protein 1..538
/product="60 KD RO PROTEIN"
Region 93..98
/region_name="Domain"
/note="RNA-BINDING (RNP2) (BY SIMILARITY)."
Region 124..131
/region_name="Domain"
/note="RNA-BINDING (RNP1) (BY SIMILARITY)."
Region 239
/region_name="Conflict"
/note="K -> R (IN REF. 2)."
Region 515..538
/region_name="Conflict"
/note="GMLDMCGFDTGALDVIRNFTLDMI -> ALQNTLLNKSF (IN REF. 2)."


ORIGIN
1 meesvnmqmp lnekqiansq dgyvwqvtdm nrlhrflcfc seggtypike qklglенаа
61 lirliedgrg ceviqeiksf sgegrttkqe pmlfalaics qcsdistkqa afkavsevc
121 ipthlftfiq fkdldkesmk cgmwgralrk aiadwynekq gmalalavtk ykqrngwshk
181 dlrlrshlqp sseglaivtk yitgwkevh elykekalsv eteklkkyle avekvkrtkd
241 elevihliee hrlvrehllt nhllkskevsk allqemplita llrnlgkmta nsvlepgmse
301 vslvceklon ekllkkarih pfhllialet yktghglrgk lkwrpdeeil kaldaafykt
361 fktvptgkr fillavdsas mngrvlgsil nastvaaamc mvvtrtekds yvvafsdemv
421 pcptvtdmtl qvvlmamsqi paggtcdslp miwaqkntnp advfivftdn etfaggvghpa
481 ialreyrkkm dipaklivcg mtsngftiad pddrgmldmc gfdtgaldvi rnftldmi

THIS PAGE BLANK (USPTO)

Figure 38A

File Edit Search Type Help

Accession Code Query



Tools

←

?

?

X

Sequences

?

?

?

? Acc
y123

Ligands

? ADP
NAD
FAD

? Phe
Ala
Glu

Select a database and enter a code:

Database	
GenBank	
Accession	<input type="radio"/>
Accession Version	<input type="radio"/>
GI-NID	<input type="radio"/>
protein id	<input type="radio"/>
GI-PID	<input type="radio"/>
SWISS-PROT	
Accession	<input checked="" type="radio"/>
ID	<input type="radio"/>
PDB	
ID	<input type="radio"/>

P10155

Submit Reset

THIS PAGE BLANK (USPTO)



THIS PAGE BLANK (USPTO)

Aligned Sequence Display

Query details:

P10155: 60 KDA RO PROTEIN (60 KDA RIBONUCLEOPROTEIN RO) (RORNP) (SJOEGREN SYNDROME TYPE A ANTIGEN (SS-A))



100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

100%

View alignment

Total selected: 1

Reverse-Maximised PSI-BLAST results

Page 1 of 1

Total hits: 13

Selected: 0

Select all

Deselect all

Cluster	Detail	Code	Title	Organism	%ID	Query/rgn	Target/rgn	Filter	E-value
o	D	AAA35532.1	Not given	<i>H.sapiens</i>	98.0	1 - 516	1 - 516	1	0.00E+00
o	D	AAE190490.1	Not given	<i>M.musculus</i>	90.0	1 - 538	1 - 537	1	0.00E+00
o	D	AAC53142.1	Not given	<i>M.musculus</i>	90.0	1 - 538	1 - 538	1	0.00E+00
o	D	AAC38001.1	ribonucleoprotein	<i>X.laevis</i>	76.0	1 - 530	1 - 530	1	0.00E+00
o	D	AAB81552.1	Not given	<i>H.sapiens</i>	100.0	1 - 188	1 - 188	1	1.00E-111
o	D	CAX98241.1	Not given	<i>C.elegans</i>	36.0	7 - 538	354 - 643	2	0.00E+00
o	D	AAF10833.1	Not given	<i>D.radiodurans</i>	34.0	7 - 538	15 - 531	2	0.00E+00
o	D	AAF45676.1	Not given	<i>D.melanogaster</i>	17.0	1 - 537	64 - 638	3	1.00E-126
o	D	CAB46406.1	hypothetical protein	<i>S.coelcolor</i>	26.0	266 - 381	273 - 372	3	6.00E-25
o	D	AAG59043.1	telomerase protein	<i>M.musculus</i>	21.0	264 - 379	513 - 627	3	6.00E-05
o	D	AAB51690.1	telomerase protein component 1	<i>R.novogicus</i>	23.0	264 - 375	517 - 627	3	2.00E-04
o	D	AAC46601.1	telomerase component p80	<i>T.thermophila</i>	27.0	264 - 329	361 - 425	3	7.00E-04
o	D	AAC51107.1	telomerase-associated protein TP-1	<i>H.sapiens</i>	25.0	264 - 327	507 - 569	3	0.00E-03

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences

Sequences



THIS PAGE BLANK (USPTO)

Figure 39

AI Eye output (January 2, 2001 1:29 PM)

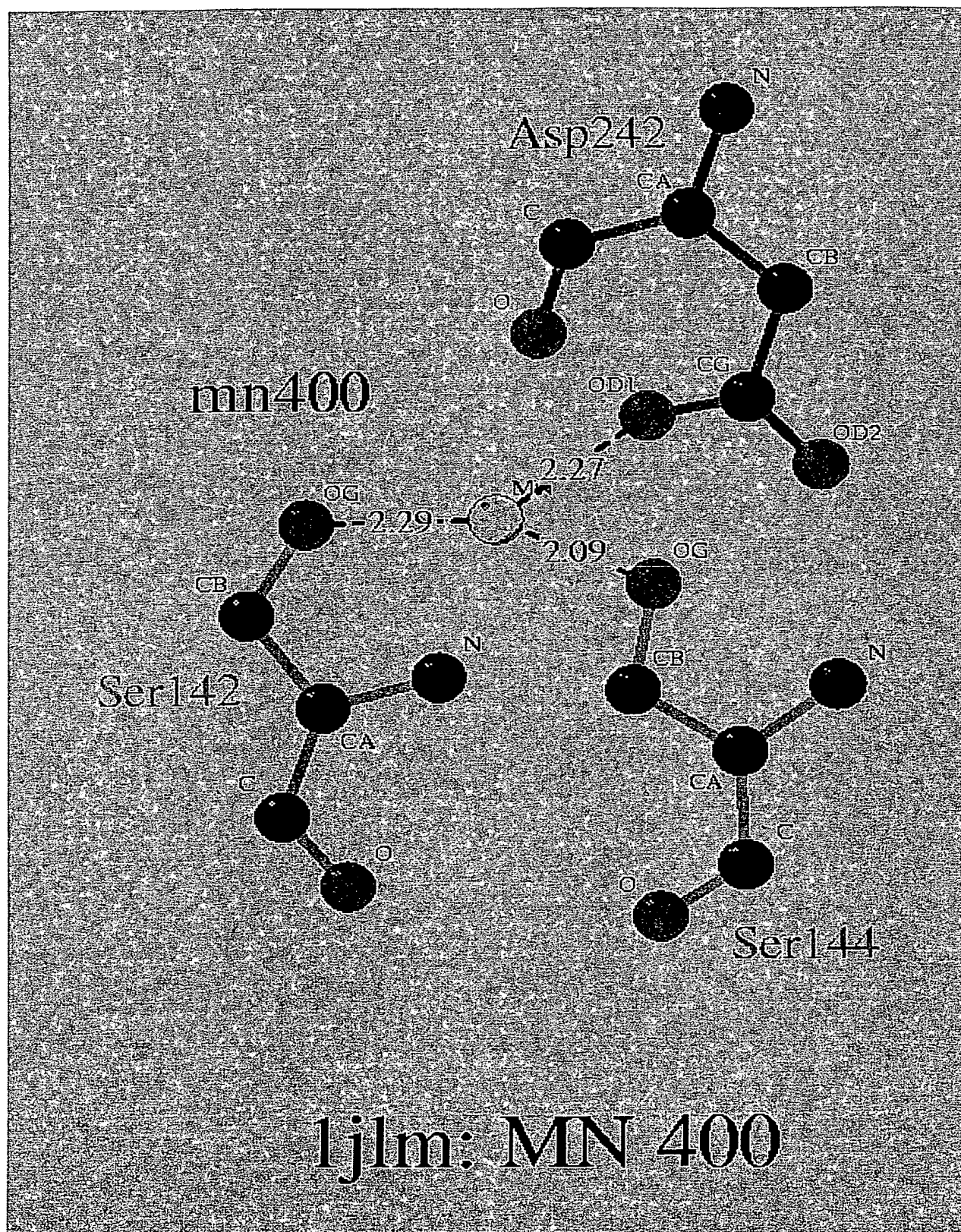
P10155 1JUN	MEESVNQMQPLNEKQIANSDQGYVWQVTONRPLRFLCFGSEGGTYI KEQKLGLENAEALIRLIEDGRGCEVIQEIKSF	10	20	30	40	50	60	70	80
P10155 1JUN	SQEGRTTKQEPMLFALACSQCSDISTKQAAFKAVSEVCRIPHTLFTFIQFKKDLKESMKCGMWGRALPKAIADWYNEKG	90	100	110	120	130	140	150	160
P10155 1JUN	GMA L A L A V T K Y K Q R N G W S H K D L L R L S H L K P S S E G L A I V T K Y I T K G W K E V H E L Y K E K A L S V E T E K L L K Y L E A V E K V K R T K D	170	180	190	200	210	220	230	240
P10155 1JUN	E L E V I H L I E E H R L V R E H L L T N H L K S K E V W K A L L Q E M P L T A L L R N L G K M T A N S V L E P G N S E V S L V C E K L C N E K L L K K A R I H	250	260	270	280	290	300	310	320
P10155 1JUN	P F H I L I A L E T Y K T G H G L R G K L K W R P D E E I L K A L D A A F Y K T F K T V E P T G K R F L - - L - - A V D V S A S M N Q R P V L G S I L N A S T V A	330	340	350	360	370	380	390	400
P10155 1JUN	A - - - - - A M C M V V T R T E K D S Y V V A F S D E M V P C P V T T O M T L Q Q V L M A M S Q - - P A G G T D C S L P M I W A Q K T N T - - - P A D V F I V F	410	420	430	440	450	460	470	480
P10155 1JUN	M E Q L K K S K T L F S L M Q Y S E E F R I H F T F K E F Q N N P N P R S L V K P I T Q L L G R T H T A T G I R K V V R E L F N I T N G A R K N A F K I L V V I	490	500	510	520	530	540	550	560
P10155 1JUN	T D N E T F A G G V H P A I A L P R E Y R K K M D I P A K L I V C G M T S N G F - - - - - T I A D P P D D R G - M L D M C G F D T G A L D V I R N F T L D M I								
P10155 1JUN	T D G E K F G D P L G Y E D V I P E - A D R E G V I R Y V I G V G D A F R S E K S R Q E L N T I A S K P P R D H V F Q V N N F E - - A L K T I Q N Q L R E K I F								

P10155
1JUN A

THIS PAGE BLANK (USPTO)

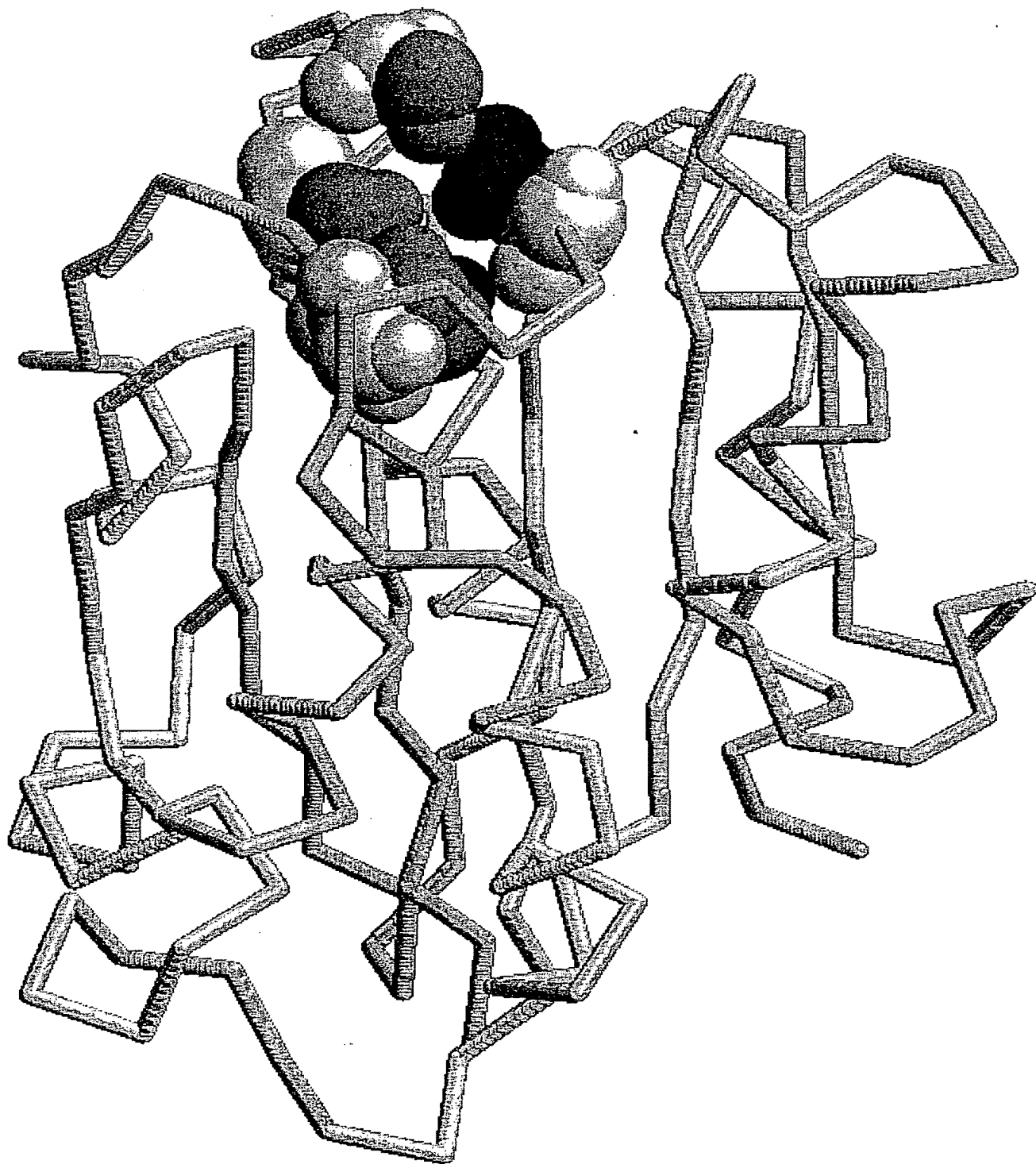
Figure 40A

009921



THIS PAGE BLANK (USPTO)

Figure 40B





THIS PAGE BLANK (USPTO)

Fig 41

```

P10155
AAF19049.1
AAC38001.1
CAA98241.1
MEESVN-----Q-----MQLNEKQIANSDQGYVWQ
MEGSAN-----Q-----LQPLSETQVNVSEGGCVWQ
MEATMD-----Q-----TQPLNEKQVPSSEGCYVWQ
-----MADELNEFQEAAGNFEEALMRLSNVCAARLRMQMLESDVEITVVDGELKRVPQMEKVKDGGQVENNAGGFVFP

P10155
AAF19049.1
AAC38001.1
CAA98241.1
VTDMMRLHRLFCFGSEGGTTYIKKQKLGLENAEALIRLIEDGRGCEVIQKSFQEGRTTKQEPMLFALAI CSQCSD
VTDMMN-VRRFLFCFGSEGGTTYIKKQKLGLENAEALIRLIEDGRGCEVIQKSFQEGRTAKQEPLLFALAVCSQCAD
VSDMMNRLRRLFCFGSEGGTTYIKKQKLGLENAEALIRLIEDGKGCEVVQKSFQEGRAAKQEPTLFALAVCSQCSD
VSDETQVRRFLILGSDKGSYHQSSEKITIDNAQRIRKIEQGNHGMVLKELALINAENRNPKNAMIFTLAICARIST

P10155
AAF19049.1
AAC38001.1
CAA98241.1
ISTKQ-----A-AFKAVSEVCRIPTHLFTFIQFKKDLKESMKCGM-----WGRALRKAIADWYN
INTKQ-----A-AFKAVPEVCRIPTHLFTFIQFKKDLKESMKCGM-----WGRALRKAVADWYN
IKTKQ-----A-AFRAVPEVCRIPTHLFTFIQFKKDLKESMKCGM-----WGRALRKAVSDWYN
HDTTKKTECPMLNAYS DYIRALHDSALDLIPEVCRTPTHLFEFVDYCQTI SESTKAGGAKSSTGWGRSMRNAISKWYT

P10155
AAF19049.1
AAC38001.1
CAA98241.1
EKGGMALALAVTKYKQRNGWSHKDRLRLSH---LKPSSSEGLA---VTKYITKG-----WKEVHEL
EKGGMALVALVTKYKQRNGWSHKDRLRLSH---LKPSSSEGLA---VTKYITKG-----WKEVHEE
TKDALNLAMAVTKYKQRNGWSHKDRLRLSH---LKPANEGLTM---VAKYVSKG-----WKEVQEA
TKTTEKLAMLLTKYPQREGWSHRDLFRLAHPNLMDSRSHGQSEDRLEREQLFRAFVKGD LVKRKRKMSVEEVAEVEKV

P10155
AAF19049.1
AAC38001.1
CAA98241.1
YKEKALSV-----ETEKLLKYLEAVEKVKRTKDKDELEVIHLIEEHRLVREHLLTNHLKSKEVWKALLQ-EMPLT
YKEKALSV-----EAEKLLKYLEAVEKVKRTKDDLEVIHLIEEHQLVREHLLTNHLKSKEVWKALLQ-EMPLT
YKEKELSP-----ETEKVLKYLEATERVKRTKDKDELEIHLIDYRLVREHLLTIHLKSKEIWKSLLO-DMPLT
WDKKALKLPYTEEQLIKKEQSRAALNLVEAYLKLKNEQSEEVIVAAIKKHGLVREHLLPTTSLNSKLVWETLFDVSMPT

P10155
AAF19049.1
AAC38001.1
CAA98241.1
ALLRNLGKMTANSVLEPGNSEVSLVCEKLCNEKLLKKARIHPFHILIALETYKTGHGLRGKWKRPDEEILKALDAAF
ALLRNLGKMTANSVLEPGNSEVSLICEKLSNEKLLKKARIHPFHVLIALETYRAGHGLRGKWKIPDKDILQALDAAF
ALLRNLGKMTADSVLAPASSEVSSVCERLTNEKLLKKARIHPFHILVALETYKKGHGNRGKLRWIPDTSIVEALDNAAF
AMIRNLAKMTVVGALDE--KRVDNIVKRLTDQEEELRRSRIHPINLLTARAVYAQGRGDKGSLTWEPNQKICDALEAGF

```

THIS PAGE BLANK (USPTO)

Page 41 cont...

2

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)